Center for Advanced Computation

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801

CAC Document No. 151

IRIS

Illinois Resource Information System
Feasibility Study
Final Report

April 30, 1972

IRIS

ILLINOIS RESOURCE INFORMATION SYSTEM

Feasibility Study

Final Report

April 30, 1972

Center for Advanced Computation
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This volume presents, in detail, the findings and recommendations of a study to determine the feasibility of an Illinois Resource Information System (IRIS). A shorter, summary document is available. An Illinois Data Catalog was also produced as part of the feasibility study.

Background

The State has several reasons for interest in a computer based geographic information system. First, the State's growing population and the growing number of services require large amounts of data for proper administration. Second, the impact of this growing population and its higher standard of living on the limited resources of air, water, and land has become a matter for serious concern. New and different types of data are required to insure optimal allocation and adequate protection of these limited resources.

The recently adopted IMPACT 70's program exemplifies the State's long experience with automatic data processing and its forward-looking approach to information programs. IMPACT 70's includes a plan for administrative data processing and proposes a mechanism for setting data standards.

The Illinois State government has also been quick to adapt itself to the changing needs of society. Shortly after the passage of the National Environmental Policy Act of 1969, the State of Illinois adopted the Environmental Protection Act of 1970. This Act created three new agencies charged with the protection of Illinois' land, water, and air resources. Two of these agencies, the Pollution Control Board and the Environmental Protection Agency, immediately set out to establish and enforce a set of regulations aimed at correcting past

mistakes and making up for years of inadequate control. The third agency, the Institute for Environmental Quality, was charged with looking beyond the remedial approach to pollution control and developing a sound environmental planning program for the future. It was this responsibility which led the Institute to confront the problem of planning information.

In May 1971, the Office of Planning and Analysis was created by Executive Order and charged with coordinating nearly all planning activities in the State. This agency was in a position to see the potential for overlap and duplication in the collection and use of planning data.

In the past, it was often sufficient for each functional planning agency to gather the data needed to plan and administer its own programs. The few types of data which are exceptions to this rule, vital statistics for example, were provided by other State agencies under specific statutory directives. Recently, with agencies often competing for limited resources, the need for coordinating planning efforts has become apparent. In addition, existing federal requirements and proposed State requirements for environmental impact studies have radically changed the data requirements of functional planning agencies. For example, an airport planner years ago needed access to little more than airport design data; he now needs access to data on a variety of other subjects, including alternate modes of transportation, in order to write an envionmental impact statement for a proposed airport. The infeasibility of placing such a diverse data collection and information processing burden on each functional planning agency is apparent; the feasibility of alternatives was not. The alternative of a statewide geographic information system was investigated in detail.

Geographic Information Systems

Geographic information systems are computer systems used for storage, retrieval and analysis of data which refer to specified geographic locations. The many different types of data needed for these programs have only one thing in common, a geographic location. For example, a population, an average income, a soil type, depth to bedrock, and the availability of ground water may all be stored in a geographic information system as attributes of a specific parcel of land such as a city block. Other types of geographic data such as air pollutant emissions may refer to points, such as the locations of the emitting smokestacks. Still other geographic data refers to networks, such as streams or roads. Examples of data which are not geographic include staff or administrative data such as payrolls, program-related data such as the status of federal programs within a particular county, etc.

The term "geographic information system" has been used to describe many different types of computer systems having a wide variety of capabilities. The majority of these systems were built to solve specific problems and could not be readily adapted to meet changing user needs.

Many geographic information systems have been built in other states, but most have failed to gain wide acceptance because of limited analytical capabilities or limited data bases.

The geographic information system under consideration is called the Illinois Resource Information System (IRIS). The proposed IRIS would contain extensive point, network, and area related data. It would allow the user to search and analyze the comprehensive geographic data base using a language much like ordinary English. The system would have extensive statistical capabilities, and would be able to prepare maps, charts, and tables.

Providing the user with analytical capabilities is very important. Given a very large data base, there is an almost infinite

number of ways to analyze, aggregate, and present this information. A comprehensive geographic information system would provide users with access to raw, basic data and the capabilities for analyzing it in real-time. This would be a marked change from the present method of transmitting aggregated data between agencies via reports and reprints.

Besides helping to solve problems of information distribution, a geographic information system would also address problems associated with the changes in data requirements for comprehensive planning. Data has traditionally been cataloged, stored, and retrieved according to subject area. While this system may have worked in the past, it does not meet the needs of today's planners. While it was once sufficient for planners to have access to all available information on a particular subject, it is now necessary to have access to all available data relating to specified geographic areas. A geographic information system is capable of presenting information in this form, and expediting the comparison and analyses of diverse sets of data. In effect, it computerizes simple analyses now done manually by overlaying maps and it permits much more sophisticated analyses of the relationships between various data elements.

The Feasibility Study

The IRIS feasibility study began in May, 1971 and was supervised by a policy committee consisting of representatives from the Institute for Environmental Quality, the Office of Planning and Analysis, and the Bureau of the Budget. The study began with a survey of potential users of a geographic information system during which officials at more than 50 state, federal, and local agencies in Illinois were interviewed. On the basis of this initial survey, the IRIS staff compiled an Illinois Data Catalog and analyzed the present status of data collection and utilization in Illinois. Chapter 2 discusses the user survey as well as several problems associated with inter-agency

data flow. Solutions to these problems are recommended in Chapter 3. The Data Catalog is published as a separate volume.

A survey of the state of the art in geographic information systems was conducted concurrently with the survey of potential users. Programs at over 60 universities, governmental agencies, professional study groups, and private companies in six different countries were investigated and are summarized in Chapter 4. Preliminary design recommendations for the Illinois Resource Information System (IRIS) are presented in Chapter 5. A 3-year program to develop and implement IRIS is recommended. The costs and schedules for this proposed program are presented in Chapter 6.

The problems and circumstances which create the need for IRIS are not unique to Illinois. Due to the program's innovative nature and the nation-wide need for its products, it is recommended that the State and the University of Illinois jointly seek funding from an appropriate federal agency. The Research Applied to National Needs (RANN) program at the National Science Foundation has already been contacted. The Ford Foundation, which sponsored the development of the Natural Resource Information System (NARIS) at the Center for Advanced Computation is also interested in the IRIS program.

CHAPTER 2

SURVEY OF USER NEEDS AND THE STATUS OF DATA COLLECTION AND UTILIZATION IN ILLINOIS

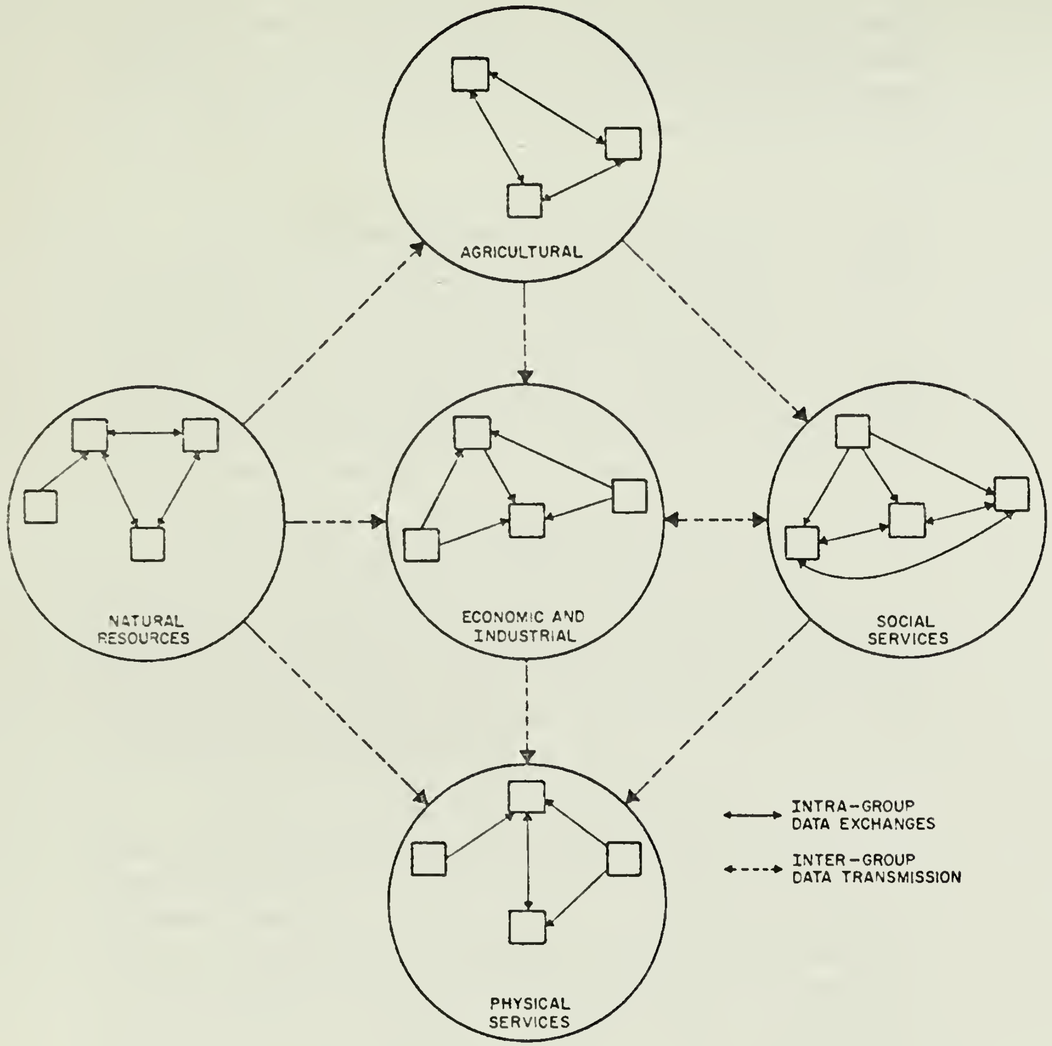
A major objective of the IRIS Feasibility study was to document the need for and the potential uses of a geographic information system for state agencies. To this end, the IRIS staff analyzed the current status of data collection and utilization in Illinois.

The first phase of the study was a broad survey of more than 50 federal, state, and local agencies in Illinois. In an initial interview, IRIS staff described the capabilities of geographic information systems to State personnel. These professionals were then asked to describe the exchange and use of geographic information pertinent to their agency. From these interviews, an Illinois Data Catalog describing nearly 200 files of geographically referenced data for the State was compiled.

Through the broad survey, functional distinctions between state agency programs became apparent. Five groups or clusters of agencies were defined according to their use and exchange of similar types of data (Figure 1). Information flow between groups is slight compared to that within each group.

Nine agencies, at least one from each group, were then selected for more extensive analyses. Several contacts were made with these agencies to determine inter-agency and intra-agency data requirements and to define existing problems of information use and exchange. At the same time, as preliminary results of the state of the art survey became available, it was possible to describe to the potential users the capabilities which could be provided in a geographic information system.

In this way, agency personnel were able to specify those system characteristics which would be of most help to them.



GROUPS OF POTENTIAL USER AGENCIES

FIGURE 1

Survey of Potential Users

In this section, the five groups of agencies identified in the broad based users survey are described, and the status of data collection and utilization within these groups is discussed.

Social Service Agencies

Agencies described as social service agencies are concerned directly with the human resources in Illinois. The information generated and used by these agencies is about people and services.

Several agencies in this group can be characterized by their involvement in single purpose programs: The Departments of Public Health, Public Aid, Mental Health, Corrections, and the Office of the Superintendent of Public Instruction. Other agencies, such as the Office of Human Resources, the Office of Planning and Analysis, and the Institute for Social Policy, play a more general planning and research role.

Those agencies that provide a single identifiable service follow a universal pattern in their data collection and use. Statutory obligations require collection and processing of a single type of information. This information is normally of high quality since a major staff effort is involved in its collection. But when data collection efforts are oriented towards only one agency's needs, their value to other agencies is diminished. This often leads to overlap when another agency has to collect similar data but in a slightly different form.

Agencies which have comprehensive social service programs must evaluate many different types of information for specific regions in the state. The information is frequently used to construct social indices to determine the kinds of social programs needed in an area. These comprehensive agencies are not directly served by any information

collecting agencies and do not have a large data collection staff. Currently their method of obtaining information for program planning involves using published documents from other agencies and making informal agreements to obtain unpublished records and special data service.

All social service agencies are concerned with the geographic distribution of the populations they serve. For example, they must locate people who need mental health care in order to provide nearby service facilities. If only coarse resolution geographic data are available (e.g. county aggregations), it is impossible to determine exactly where within a county a facility should be located. In large urban areas, it would be desirable to have information available at very fine geographic resolutions, such as the census tract or block level.

Census data: Data from the decennial census of population and housing and other periodic censuses and surveys are a valuable source of information for social service agencies. Unfortunately, the time lag between census counts and publications greatly decreases the utility of the information. Since the surveys are usually repeated only once every five or ten years, agencies which rely on this data require other information sources in the interim.

In order to provide current population data, the Department of Public Health is statutorily responsible for collecting and disseminating monthly vital statistics for the State of Illinois, including counts of births, deaths, fetal deaths, marriages, and divorces. Official monthly estimates of the total population are compiled from this data and used in nearly raw form for all official per capita statistics in Illinois. Useful demographic information is associated with vital statistics data, but it is not transmitted as regularly as the total population counts. Nor is there a mechanism for synthesizing various requests for such information into an official data transmission effort. General distribution of all potentially useful

data is infeasible because of bulk and a limited audience. For these reasons, collected data are rarely used outside the Department of Public Health. Demographic data produced by other agencies are subject to similar problems.

Locating service recipients: Individuals receiving public aid, mental health care, or welfare, for example, frequently qualify for aid from other state and federal programs. However, locating these individuals is difficult. Although public schools receive reimbursement for services rendered to children in the Aid to Dependent Children program, each school district must determine for itself the number of ADC children living within its boundaries.

This is a typical problem among agencies which have developed their own ways of geographically referencing and storing data. Inevitable inconsistencies make it difficult for one agency to reaggregate another's data for its own purposes. Thus, user agencies must make special demands on collecting agencies or do without the information.

Detailed social indices: The Department of Labor and the Institute for Social Policy in the Department of Public Aid are evaluating the Emergency Employment Act in Illinois, a process which requires controlled access to confidential files maintained by the Illinois State Employment Service. Through a computer system, the Institute retrieves detailed information about the demographic characteristics of Emergency Employment Act job applicants. The computer system permits evaluation without revealing the names or the identification numbers of job applicants, thus maintaining the confidential nature of the information.

The evaluation will determine relationships between job success and applicant characteristics such as welfare status, experience, and sex. This requires computerized manipulation of basic data to calculate

many statistical indices which cannot be defined in advance and requires access to raw confidential data. The study, more detailed than most program evaluation and planning activities, is possible only as a result of close cooperation between the Institute for Social Policy and the Department of Labor.

Economic and Industrial Agencies

Agencies in this group are concerned with every aspect of the State economy concerning jobs and income, and the institutions and activities which produce them. Individual agencies in this planning area are the Department of Business and Economic Development, the Department of Labor, the Department of Revenue, and the Bureau of the Budget. Much of the data used by these agencies comes from federal sources, the Department of Commerce in particular.

Programs are closely related to those in the social service area, but with some major distinctions. Because economic aspects of the population are considered, the focus of information is away from individuals and toward economic profiles of geographic areas. Services rendered relate only indirectly to individual needs through maintenance of public funds by taxation and through maintenance of a high employment rate. Much of the program planning requires interaction with the private sector to encourage private investment in Illinois. The regional economy may influence an area's growth, types of jobs, income levels, amount of personal migration, and other factors. Therefore, economic indicators are calculated to help define growth trends and determine an area's need for public facilities and physical services. The particular mission of a given agency defines the type of information for which it is responsible. Within each agency, data are kept on file in a form best suited to that agency's needs. Unfortunately, agencies often have insufficient time and staff to reprocess this data to meet specific requests from other potential users.

Data for tax reform studies: In order to evaluate new taxation strategies, the Bureau of the Budget needs access to detailed income data. The evaluation is made by determining the impact of a new tax policy on wage earners and on the tax revenue produced in various regions of the state. This requires a large amount of information about income, property, sales, and other taxable items.

The information must be available at the community level in order that the effect of a statewide tax policy on individual communities and their inhabitants can be assessed. The distribution of personal income among population sub-groups in an area must also be measured. When this information is available community by community, it will be possible to predict overall impacts that would result from alternative tax strategies.

Often it is necessary to examine confidential data in order to determine the impact of a given taxation strategy. For example, it may be necessary to examine the income levels and property tax rates in a group of neighborhoods in order to equalize the tax burden. Such confidential data is presently published only in highly aggregate form for large geographic areas. A means is needed for providing this data for comparison and analyses in smaller areas while still protecting its confidentiality.

Economic profiles: A detailed economic profile of a region is one of the most basic and useful indicators for both comprehensive and single function planning. The Federal Departments of Commerce and Labor have provided most economic data. Although this information is very detailed, it is usually updated only every five or ten years, with publication dates often lagging several years behind the data collection.

The Department of Business and Economic Development has responded to this lack of current socio-economic data for the State. It

publishes annually Illinois Regional Economic Data Book which brings together federal, state, and county records from many available sources, and presents economic profiles for 16 multi-county regions of the State. The Data Book is intended primarily as a reference for individuals planning to invest in an Illinois business or industrial plant; unfortunately some agencies, particularly the social agencies, find it inadequate for their uses. It has insufficient geographic resolution and lacks income distribution breakdowns. Data of fine geographic resolution are needed because the proximity of the public to job markets, consumer products, and other public services does not normally coincide with the boundaries of these multi-county regions. For example, if a firm was considering locating a plant near the boundary of one of these regions, it would attract workers from two or three multi-county regions. Aggregated employment and work force data would be virtually useless. Therefore, fine resolution data is necessary.

Much of the basic data used to compile the Data Book could also be used to prepare fine resolution profiles. More detailed economic profiles could be cross referenced to the population characteristics and personnel income information used by social service planners. This type of data sharing now occurs only when there is cooperation on a specific project, such as that described above between the Institute for Social Policy and the Department of Labor.

The needs of the social service and economic agencies are similar. Both need a formal mechanism to facilitate information sharing. Once the information transmission requirements have been identified, a mechanism for transmitting confidential data in a secure manner must be provided. Finally, the agencies will need a means for comparing and analyzing many different types of data over geographic areas or population subgroups.

Physical Service Agencies

In the past, most agencies engaged in providing specific physical services have operated in a manner similar to the social service agencies. They have been mission oriented, with a specific charge to provide facilities such as highways and waterways, and services such as drainage, soil and water conservation assistance, etc. They have also determined the need for services and planned for construction of facilities.

In recent years, several comprehensive planning agencies have been established to coordinate mission-oriented agency operations. The State Office of Planning and Analysis and regional planning agencies exercise some degree of control over physical services provided not only by the government, but also by the private sector. Since they have broad responsibilities, comprehensive planning agencies must rely heavily on data produced by others. They are faced with serious problems due to inconsistencies in data obtained from a variety of suppliers. General socio-economic data are usually obtained from economic and social agency publications. Detailed physical data needed for planning and providing physical services are obtained through special internal data collection efforts. More general physical data are obtained from natural resource agencies.

Mission oriented planning: In the past, most mission oriented agencies have been able to provide and plan for specific physical services using a relatively limited data base. Recently, however, the increased pressure on finite natural resources and laws such as the National Environmental Policy Act of 1969 have forced these agencies to consider a much broader range of data in order to evaluate social and environmental impacts, and evaluate alternatives in project location and scope. These new requirements have created serious data problems in several agencies. Special collection efforts for specific projects cannot provide all the necessary data without costly overlap. Redundancy could be avoided if data were collected in a form useful to other agencies for project planning.

<u>Comprehensive planning</u>: State, regional, and local planning agencies need access to all types of data. Most of the information they use is collected by other agencies for some purpose other than regional comprehensive planning. Therefore, problems of inadequate content and inconsistent format must be overcome. Some mechanism for transmitting data requirements to collecting agencies is also required.

Often the region for which a comprehensive plan must be developed is different from that within which data are originally collected. To develop a comprehensive plan for a river basin, for example, it is necessary to obtain basic socio-economic and natural resource data and aggregate it to a new geographic area, the watershed. A geographic referencing scheme consistent among all major data types must be developed, and facilities for reaggregating and overlaying such information must be provided.

Comprehensive planning agencies are often faced with resource allocation problems. In order to allocate certain areas of land for various uses in an optimal manner, cluster analyses or weighting function analyses are often performed. Cluster analyses, used to find groups of parcels of land having similar characteristics, reduce the amount of information that must be examined in making land use decisions. In weighting function analyses, all data attributes of each parcel of land are assigned subjective ratings according to their suitability for a particular land use. These ratings are then summed to obtain a total score indicating overall suitability. Cluster analyses are impractical without the aid of a computer, and weighting function analyses are prohibitively expensive when done manually. The Northeastern Illinois Planning Commission recently employed a manual weighting function analysis to develop an open space plan. It found the analysis so expensive that only one alternative could be developed for presentation to policy makers. The Commission has expressed a strong need for a comprehensive geographic information system which would allow them to perform similar analyses inexpensively and quickly. They could

then evaluate many alternatives, permitting input from and interaction with policy makers.

<u>Natural Resource Agencies</u>

Natural resource agencies include the State Water Survey, the Natural History Survey, the Geological Survey, the Department of Conservation, the Environmental Protection Agency, and the data collection divisions of state and federal physical service agencies.

These agencies collect information describing the natural environment and interpret it for various purposes. In the past, because the economic value of natural resources was of primary interest, most measurements described a resource and indicated its potential for economic development. But recently, due to the increasing pressure on these finite resources, society's interest in environmental data has changed from a concern for quantity to a concern for quality.

Physical service agencies which use natural resource data are often only indirectly linked to the collection agencies. In many cases, formal channels do not exist for transmitting user requirements and specifications to data collectors. Data needs have changed dramatically due to various federal and state laws which require the submission of detailed environmental impact statements as a prerequisite for funding. The lack of direct links between agencies has made response to these changing needs difficult.

<u>The need for new interpretations</u>: Since the passage of the National Environmental Policy Act, new interpretations of natural resource data have been required. Very little existing information about natural resources is in the form required to measure environmental quality. For this reason, it has been difficult for physical planning groups to determine the impact of their activities on environmental quality.

Traditional data collection efforts in the geological area emphasized mineral development, topographic mapping, and stream gaging. Recently, the geological constraints on the development of land areas have been recognized and demands have been made for new types of data to be interpreted at a finer geographic resolution to show suitabilities of areas for various types of land use activities. The geographically referenced data must then be compared with soils data and other kinds of natural resource data to form a composite picture. This concern with geological constraints is typical of the growing recognition of physical limitations on the use of all natural resources.

Water quality management planning: Many federal agencies require comprehensive plans for public programs. Funds are available from the Federal Environmental Protection Administration, for example, to help improve and maintain water quality through sewage treatment plant construction. The State of Illinois is required to prepare a comprehensive water quality management plan for each major river basin in order to obtain these matching funds.

Planning requirements have created a need for more extensive data collection efforts in many areas, and for new types of environmental data. Natural resource data must be compared with social and economic data within the confines of a naturally defined area such as a watershed. The objective is to determine the effects of areawide plans on the water resource and the effects of various water quality management plans on the region.

Most water data is geographically located by reference to the stream network. It is necessary to compare this network-oriented data with other area-related data. Simultaneous analysis of these large quantities of area and network data is not possible without a comprehensive geographic system.

In summary, requirements for information about natural resources have changed radically in recent years. As a result, data collection and interpretation procedures must be changed and more formal channels for communicating data needs must be established. The data must be geographically referenced so that it can be compared to socio-economic data.

Agricultural Agencies

The traditional objectives of agencies in the agricultural sector, somewhat disassociated from other planning areas of state government, have focused on the production of food and fibrous materials and the maintainance of soil and water resources for these functions. Most data collection and planning programs are administered by the U.S. Department of Agriculture and the Illinois Department of Agriculture.

The extensive personal orientation of its planning programs makes the agricultural sector unique. Every county in Illinois has several agency district offices where staff members work closely with farmers and residents of rural communities to determine planning needs and develop rural natural resources. Other agencies have begun to recognize and need information collected by agricultural agencies. As these needs grow, data collection efforts must be modified.

Urban development: The extensive growth of urban areas has increased the need for basic information regarding the soil and drainage conditions which limit land and water development. Soil scientists have defined soil suitabilities for uses such as septic fields or landfill sites, through consideration of such factors as soil type, slope, and erosion potential. Maps showing soil limitations on various land uses have recently been published.

By amending the Soil and Water Conservation Districts Law, the State has acknowledged the important role that agricultural information

plays in land use decisions. Soil and Water Conservation Districts must now provide all natural resource data required by counties and municipalities for zoning or sub-dividing vacant or agricultural lands. The natural resource information system (NARIS), developed by the Center for Advanced Computation for the Northeast Illinois Natural Resource Service Center, provides this capability.

Rural development: Much socio-economic information about farmers in rural communities is collected by the Division of Agricultural Statistics of the Illinois Department of Agriculture. Several reports published by the Cooperative Crop Reporting Service emphasize agricultural production information. Extensive socio-economic data is obtained in the annual farm census. This information could be geographically correlated with production data and the economic and social information produced by other agencies to form a comprehensive data source for rural economic development programs.

Events such as the amendment to the Soil and Water Conservation Districts Law are an indication of new forms of inter-agency cooperation planned in the Agricultural sector. This cooperation should make it possible for agencies to increase the value of their data by geographically referencing it in such a way that it can be compared to other natural resource and socio-economic data. A comprehensive geographic information system would facilitate these comparisons.

Client Analysis and System Specifications

From the large group of potential users initially interviewed, nine agencies were selected for a more detailed analysis of their information requirements. This group of "client agencies" included representatives from each of the major clusters of agencies identified in the user survey. These client agencies were selected to form a representative nucleus of information users from which a larger community of IRIS users could develop.

From the social service agencies, the Department of Mental Health, the Department of Public Health, and the Department of Public Aid were selected for further study. The Division of Highways of the Department of Transportation was selected as an agency responsible for planning and providing a specific physical service. Three planning agencies, the State Office of Planning and Analysis, the Northeastern Illinois Planning Commission and the Southwestern Illinois Metropolitan Area Planning Commission, were also selected from the physical service group. The Department of Business and Economic Development was included because it assembles, analyses, and disseminates economic, employment, and financial data for counties and municipalities throughout the State. The Environmental Protection Agency, which with the Institute for Environmental Quality shares responsibility for the development of water and air quality management plans, was selected because of its obvious need for a variety of geographically referenced data. The Northeast Illinois Natural Resource Service Center (NRSC) represented the agricultural sector.

Information flow problems:

As a result of the more detailed analysis of information flow through these and other agencies, three basic problem areas were identified. First, it was recognized that large amounts of overlapping data existed due to the nature of the present administrative structures.

Agencies often justify minimal data collection efforts financed by earmarked funds allocated to specific functional areas. It is possible to have three different agencies sampling the water quality at a specific point for three separate reasons: the U.S. Geological Survey for general water quality monitoring, the Environmental Protection Agency for drinking water suitability, and the Department of Public Health for bathing beach suitability, for example. Each agency might measure concentrations of 20 parameters, of which 15 might be common to all three analyses. A simple lack of communication and coordination prevents the possibility of one agency sampling for 27 parameters, thus meeting the needs of all three agencies. If there were readily available administrative channels to transmit data requirements to data collection agencies, and if provisions were made for the requesting agency to bear the appropriate incremental cost, such situations could be avoided.

Second, there is no central clearinghouse for authorizing data collection activities. The establishment of a common repository for basic, geographically referenced data would provide the means for monitoring data collection activities and for eliminating overlap and duplication.

Third, the content of data now transmitted between agencies was found to be inadequate for many purposes. As raw data is processed, aggregated, and summarized for publication by one agency, its value to most other agencies decreases significantly. The value of this raw, geographically referenced data is presently lost due to formatting inconsistencies and to the lack of a computer system to overcome the logistical problems associated with the bulkiness of such data.

Need for analytical capabilities:

Analyses of client needs led to the identification of the above data specification and transmission problems. Changes needed to

alleviate these problems require the support of a large, online computing facility. Such a computer system is required for the agencies to analyze large quantities of geographically referenced data.

The need for analytical capabilities could not be ascertained during the initial phases of the user survey; they became evident to IRIS project personnel during the detailed study phase. Since a diverse set of users was represented, needs were expressed for a variety of analytical capabilities.

There are two ways to provide the analytical capabilities needed by this large user community: with a standard computer system via a middleman, or with a sophisticated computer system via direct user interface. In the first case, each user could define a particular problem to a computer programmer who would then write a program to manipulate the appropriate data items and perform the requested analyses. The logistical problems of this alternative are obvious, but the method has the advantage of requiring a simpler and less expensive computer system. Whether this savings would offset the cost of retaining programmers would depend on individual situations. Alternatively, users could be provided the required analytical capabilities by designing a computer system with several simple English language interfaces for different types of users to perform analyses directly, in a conversational mode. Direct interfaces have been successfully developed for users in this country and Canada. Since this type of interface encourages direct, unimpeded data analysis, it was decided to consider the possibility of developing IRIS to provide not only online data retrieval capabilities, but also conversational access to the data base through user oriented analysis languages.

Computer system recommendations:

The computer system design was influenced strongly by the needs for analytical capabilities expressed by the IRIS user community.

Several potential users requested similar analytical capabilities, but used different methods to perform the same analytical operations. The regional planner looking at a geographic region described the region by specifying a set of land parcels having certain characteristics. The public health official, looking at the relationship between air quality and certain demographic and health characteristics, specified a certain level of air quality and defined a sample population group for study. In both cases, the analyst specified an information key inside the computer; the first called it a region, the second called it a sample. Client analyses led to the conclusion that no single user interface language could be designed to accommodate the variety of users to be served. Therefore, the feasibility of designing separate interface packages for several distinct classes of users was investigated in detail. As a result, two user interface packages are recommended for the initial system. The first will be an expanded version of NARIS, containing more powerful analysis capabilities and a more flexible user language to accommodate users in the physical service, natural resource, and agricultural sectors. The second interface package, written for users in the economics and social service sectors, will provide capabilities for data management and statistical analyses. Other general capabilities, needed by most users, were also identified and are described in detail in Chapter 5.

Initial data base recommendations:

As a result of both the broad based user survey and the detailed analyses of client needs, two initial data bases are recommended. A statewide data base, consisting mainly of existing data at a variety of geographic resolutions, was selected to meet many of the needs of the client agencies. The recommendations are designed to encourage other agencies to contribute additional data so they can compare it with information already in the system. Because each data file added to the system by any agency will increase the value of the

IRIS data base to other users, the data base is designed to grow, and the computer system will be designed to accommodate this growth.

A fine resolution data base should be developed for a smaller geographic area within the State. Its purpose is twofold. It will provide planners and researchers with a combination of high quality natural resource and socio-economic data and the capabilities to manipulate it using IRIS. The availability of this data for experimental purposes will enable users to determine the cost-effectiveness of collecting different types of data at various geographic resolutions. By working with high resolution data, at the census tract and quarter-quarter section level, users can estimate the benefits of this data and can compare them with the costs of collection.

A detailed discussion of both recommended data bases is contained in Appendix A.

The statewide data base is designed to provide the "initial investment" necessary to assure the acceptance and successful implementation of IRIS. The point can be illustrated by the following analogy. During the first decade after its invention, the purchase of an automobile was not cost-effective for most people. It took a large public investment in paved highways to make the automobile a direct substitute for the horse. Until that public investment was made, a car was only a luxury owned in addition to a horse. Afterwards, however, the benefits to potential owners exceeded the cost of acquisition. In the IRIS feasibility study, user needs dictated system specifications. But such a system will be a luxury unless a public investment is made in an initial data base. Only then will subsequent additions to the data base be self-justified, and only then will the system become a viable substitute for the presently inadequate method of transmitting information via published agency reports.

CHAPTER 3

THE INFORMATION COMMUNITY: A SOLUTION TO THE PROBLEM OF INFORMATION EXCHANGE

The heart of the information problem in Illinois is not only that there are shortages of important data, but that it is often extremely difficult for decision makers to discover what information is necessary and what is available. If anything, administrators have extremely fast access to more incomplete, overlapping, and marginally useful data than they can comprehend. What they lack are sufficient supplies of data relevant to current problems, obtainable at reasonable effort, in comprehensible form, and of tolerable reliability. A major cause of the information problem is poor distribution of data from source to users and of poor transmission of data requirements from users to sources. One consequence of this situation is that some data is duplicated while other needed information is never collected.

Planning Advocacy and the Adversary Process

Despite coordinating agencies such as the Office of Planning and Analysis, and regional planning agencies such as SWIMAPC and NIPC, the detailed work and implementation of state planning is done and will continue to be done by line agencies. Detailed planning and administration is so complicated a process that no overall agency could maintain the expertise necessary for detailed development of a long-range social, economic, or natural resource plan. Indeed, much of the required expertise can come only from day-to-day involvement with and responsibility for a particular aspect of state government. A regional planning authority may specify that a particular carrying capacity be developed in a road network, and may specify where the main traffic corridor should be located. However, the details of road design,

construction and maintenance will probably be left to the highway engineers in the Department of Transportation. In many cases the regional planning agency may find it best first to produce a general overall plan and then invite proposals from each of the line agencies involved in implementing specific parts of that plan.

It is important that healthy adversary relationships be maintained between State agencies in the development of long-range plans; only in this way can mutually exclusive positions be adequately represented. An adversary system works best when there are explicit limits on the extent to which parties are adversaries; the collaborative elements of agency relations in the planning process are as important as the adversary elements.

For example, attorneys for the parties to a civil or criminal proceeding are adversaries within the context of the particular case, but are colleagues as officers of the court and members of the legal profession. A shared set of values and professional standards allows each party to have confidence in the other's abiding by explicit and implicit rules which make the adversary procedure function in deciding between the two arguments.

The decision-making process works most efficiently when the adversary relationship is confined only to those issues where it is needed; a collaborative relationship should be established in all other matters of joint interest. One obvious area for collaboration among agencies is the collection and analysis of data for making policy decisions. Data available to all parties stands as public evidence from which the best argument is decided. Adversaries, therefore, must have a joint interest in the data being as accurate and complete as possible, regardless of the position each advocates.

Several advantages result from adversaries' having a joint interest in information collection and processing. First, a costly

duplication of effort is avoided, since, once collected, shared data must be obtained from a common archive. Second, information users and producers can be organized in such a way that it will be to the long term advantage of each to contribute information of the highest quality to the common pool. Since information placed in the system by one user is available to all, data are open to constant scrutiny from a variety of viewpoints and interests. These considerations are especially true when an agency is providing data to a group of other users of the system on a continuing basis.

The creation of a data base independent of particular cases argued by particular pairs of agencies requires that impartial, feasible, and fair standards be established for data collection, documentation, and analysis. It is a better policy for agencies to provide data to each other on a continuing basis according to a set of shared ground rules, than for an agency to be ordered to disclose particular information about a particular issue to a particular adversary. The latter approach is hardly likely to bring about enthusiastic interagency cooperation. The goal of establishing a common information base for state agencies is to exclude the acquisition and processing of undisclosed information from the adversary process. Policy decisions can then be made on the basis of what all parties must stipulate is the most reliable available data.

<u>The Uses of Information</u>

Planning and administrative information can be classified both by subject and by the uses made of it. A useful way of classifying data files maintained by the State was by the general function they served, regardless of their particular content or the agency which maintained them. Information used by state government can be classified as "statutory", "line", "staff", and "performance evaluation" data.

A limited quantity of information is produced under statutory authority by agencies whose primary mission includes information collection and dissemination. For example, the vital statistics of the State of Illinois are produced by the Department of Public Health and are used in almost raw form by agencies throughout State government. Relatively few types of data are collected and disseminated under specific statutory authority, since each new requirement implies the creation of a new agency, either in name or in fact.

Information of primary interest to an agency is that which it collects to inform itself of its own performance. Data relevant to line, mission-oriented functions is usually in the form of an inventory for those agencies which are concerned with physical resources, or of a ledger or diary for those agencies primarily concerned with the rendering of services. For example, the Outpatient Fiscal Master File of the Department of Mental Health is a ledger file containing a record of each service rendered to an individual during a single series of mental health treatments. The Outpatient Fiscal Master provides the Department of Mental Health with a record of current performance and provides the state at large with the data for computing mental health morbidity.

The Division of Highways' Roadway file contains an inventory of all roads, road boundaries, and markers in the state, constituting the State's primary source of information on roads. At present, the Roadway file is completely accessed only once each year, to distribute motor fuel taxes and federal highway funds. Despite its wealth of information, the file is not generally used because it is too massive to handle with available data processing resources. One way of obtaining more powerful tools for handling such a file is to make potential users aware of the information contained, thus giving it a wider constituency.

Line data files have a wide range of interested users, both within and outside the agencies which generate them. Once confidentiality requirements are satisfied, agencies are usually willing to distribute summaries of their line data, because such distribution supports their organizational mission, and because the nature of the data conclusively identifies and reflects credit on the providing agency.

Much data maintained by a State agency is used within the agency itself, is reasonably confidential, and is of little interest to outside users. What might be called "staff" data -- payroll ledgers, voucher files, physical property files -- is of direct concern only to staff support sections within an agency. It is disseminated only for auditing and budgetary purposes.

Finally, agencies collect and maintain data for the evaluation of their own performance, as a guide for meting out rewards, establishing priorities, and supporting budget requests submitted to higher administration. While agencies may at times be quite willing to disseminate performance evaluation data, such willingness cannot be assumed at all times; even when public disclosure is required by statute, it is impossible to legislate enthusiastic public disclosure.

The classifying of policy and planning data into the four functional categories of statutory, line, staff, and performance evaluation, leads to an important conclusion: Information which is of interest to the widest community of users is also the information which is least confidential.

Statutory data, the most widely used class of information is the opposite of confidential; its collection and publication are required by law. Line data (in aggregated form) is not restricted because it is confidential, but because there is no adequate mechanism for requesting and distributing it. Staff data, which is often truly

confidential, is also of little interest to anyone outside the agency which produces it.

Performance evaluation data presents a mixed picture. Often it is supposed to be a matter of public record, but an individual administrator's willingness to release such information obviously depends on what it has to say about his agency. A mechanism for making disclosable performance evaluation data available to a wide audience might cause occasional discomfort to public officials, but would be a powerful influence toward accountability in government.

<u>Information Exchange Between Agencies</u>

At present, State agencies prepare most of their information in whatever form and level of aggregation best fits their internal needs. They also use this readily available internal form to provide other agencies and the public with reports of aggregated data.

It is possible to group departments of the State government into five information exchange clusters. Within each cluster, there is a relatively high rate of information exchange, but between clusters, relatively little information flows. There are five major clusters of agencies exchanging information in the state of Illinois:

> Physical services data on land use, demography, housing, and transportation are maintained (and, to some extent, exchanged) by agencies such as the Office of Local Government Affairs, the Office of Planning and Analysis, and the Departments of Transportation and Aeronautics.
>
> Social service data on mortality and morbidity, mental health, education, and welfare are exchanged by service agencies such as the Departments of Public Health, Public Aid,

Corrections, and Mental Health, and by the Office of the Superintendent of Public Instruction.

Area economic data on industry, labor, income, and taxes are exchanged by the Department of Business and Economic Development, the Department of Labor, the Illinois Commerce Commission, and the Department of Revenue.

Rural economic and agricultural data on farms, soils, and drainage are exchanged by the Department of Agriculture, the Agricultural Experiment Station, and federal agencies such as the Soil Conservation Service, the Cooperative Crop Reporting Service, and the Agricultural Stabilization and Conservation Service.

Natural resource data on air, water, and land characteristics and use are exchanged by the Department of Conservation, the Geological, Water, and Natural History Surveys, the Environmental Protection Agency, and a number of federal agencies.

Pervading each functional area are the basic vital statistics data from which population estimates and projections are prepared. Almost all agencies would like instant access to population estimates and projections for the areas they serve. Many types of data simply cannot be interpreted when they are in the form of raw frequency counts. They must be transformed into per-capita counts to be of use to decision makers.

The extreme disparities of information exchange rates within and between clusters is due to similarity of agency mission and to similarity in outlook and interests of agency personnel within a cluster. Resource administrators and highway planners both are involved with topographic and geologic data and with studying and changing the

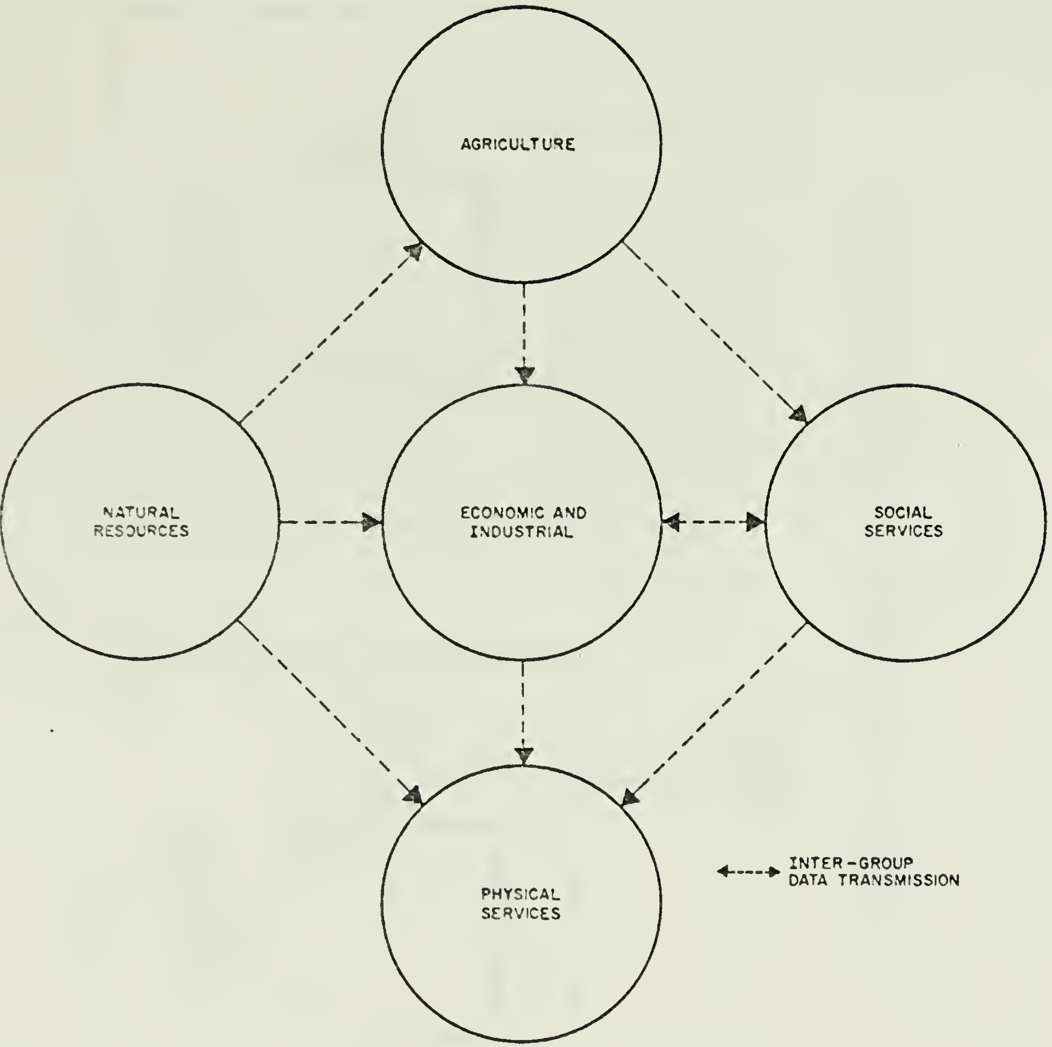
face of the land; educators and public health administrators are used to dealing with data about services rendered to individuals, and with statistics on mobility, population density, and urban facilities; water resource planners and transportation planners have had similar experiences with data about topography and land use. Where a pressing need is felt, agency boundaries are crossed. However, the need for information from other agencies goes unrecognized. Figure 1 shows the present pattern of information flow between each of the five clusters of agencies.

Differences in information terminology have resulted not only in unshared information but also in data production which, while not duplication, is largely overlapping. For example, data on wells are maintained by the State Geological Survey, the United States Geological Survey, and the State Water Survey in three completely separate files. Not only is there duplication of information on some wells and omission of others, but the files are maintained in formats which make it difficult to compare data from one agency with data from another.

Structure of an Information Community

IRIS is designed to provide administrators, planners, businessmen, legislators, and private citizens in Illinois with a tool for searching and analyzing a comprehensive library of geographically referenced information. A successful IRIS requires collaboration between users in expressing their needs and information producers in supplying IRIS with the data its users need. The operation of IRIS will require the interaction of many agencies acting as both collaborators and adversaries, as both impartial professionals and concerned advocates.

A successful IRIS management plan should facilitate each user's participation by allowing all users the greatest possible access to



INFORMATION EXCHANGE AMONG GROUPS OF AGENCIES

FIGURE 1

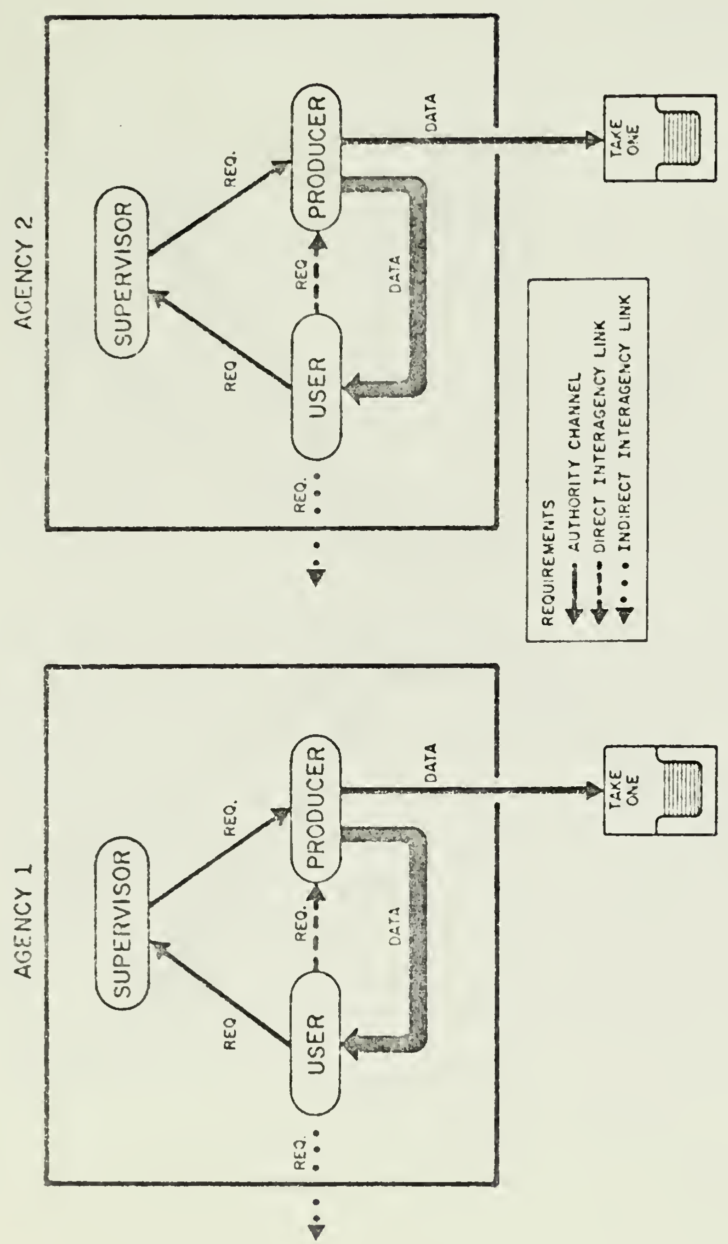
information and tools for analysis. Where each user can obtain access to as much of the available data as he feels is necessary, more efficient use will be made of and better plans and policies should emerge from the State's data base.

The best way to facilitate both cooperation and advocacy among State agencies is to require them to collaborate in the production and dissemination of information. State agencies should be organized into a community of registered information providers and users.

Some data, such as vital statistics, are produced for users throughout the State in a generally usable format. But, the largest portion of planning information is collected by State agencies for their own use and only incidentally for other state, federal, and private use.

Figure 2 illustrates the present situation. When the information provider and the information user are members of the same administrative organization, a feedback loop is established. Information flows from provider to user and information requirements flow from user to provider. When information is provided in insufficient quantity or quality, or in inappropriate form, the user can attempt to correct the situation. Since he is a colleague of the producer, the user can directly request a change in procedures. If an informal or formal request does not result in the production of satisfactory data, the user can always appeal to a common superior. The superior can then order a satisfactory change if one is technically feasible.

The situation is quite different when the information producer and user reside in different agencies. In most cases, an agency makes its information generally available either by producing extra copies of an internal report, by producing a special report for general distribution, or by responding to individual information requests. In the first two cases, the information is almost always distributed in a



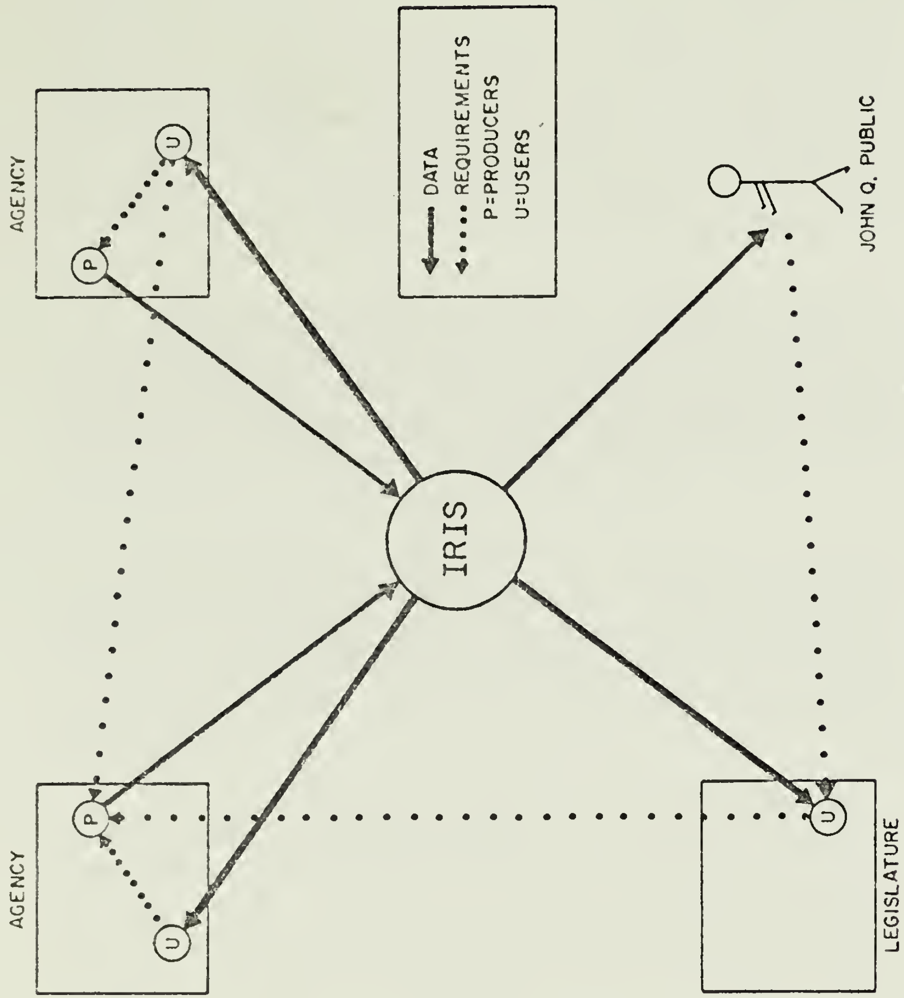
CURRENT FLOW OF DATA AND DATA REQUIREMENTS

FIGURE 2

fixed and aggregated form which the user must either accept or reject. Individual service to information demands, while providing satisfactory data, are expensive, time-consuming, and probably not supportable in the long run unless continuing budgetary arrangements are made. Thus, users of information outside the producing agency have little power to affect the quantity or quality of data available, except where they have been granted specific statutory or budgetary powers. Without such powers they must simply take what is available or develop new information collection and management efforts, incurring whatever expense is necessary.

The present situation of data duplication, anonymity, and adhocracy can be remedied only by organizing State agencies into an information community, either through legislation or by executive order. The information community should consist of those agencies in State government whose missions include the collection, processing, and dissemination of planning and policy data. In addition, the community should be composed of registered and non-registered information users. Information should be produced, not just to the standard of the agency in which it originates, but to a general standard capable of meeting the needs of most or all registered users. A non-registered user is one whose information requirements are completely met by available data, or one who has no need for active participation in the setting of information standards or requirements.

The information community will maintain channels by which information flows from producer to user, and channels by which user requirements for information flow to the information producers. Figure 3 illustrates data flow in an information community. Some data and information requests will still remain entirely within each of the agencies, but the majority of data will flow through the common geographic information system even when the data is being used in the agency in which it originated. The Legislature should be a registered user of geographic information, both to meet its own needs and to serve



THE INFORMATION COMMUNITY

FIGURE 3

as a registered representative of the public. The citizen is included in the network either as a private individual or as a businessman. He can make his special needs for geographic information known through his legislator.

An information community must have the resources to make technical decisions and to set policies. The exact form, composition, and powers of a policy making body for the information community must be decided by those in State government with the requisite experience and authority.

Overall standards for information formats, documentation, quantity, and quality should be supervised by a working group of technical representatives from member agencies of the information community. The working group will need the help of the Office of Planning and Analysis and the Management Information Division of the Department of Finance.

Each set of IRIS clients who collect or use data of a particular type should be asked to form an information exchange group. Sharing of data will require common data standards and formats which can best be set by groups of concerned clients. In order to make best use of IRIS, groups will have to set mutually acceptable data formats, determine a division of labor for data collection, and manage conflicts concerning information collection or use.

<u>Establishment and Administration of an Information Community</u>

A substantial and integral part of the system development effort should be the formation of new information exchanges and new interagency relationships where necessary. Actual and potential conflicts about data must be anticipated and managed so as to facilitate cooperation between users.

Information exchange groups will probably need both technical and organizational consulting assistance. The groups must develop feasible and acceptable data standards, and may develop coordinated data collection plans. Exchange group members will probably have diverse expectations, interests, and technical capabilities. Consultants can help the groups define and accomplish their tasks by familiarizing them with the data requirements, and the capabilities and limitations of geographic information systems. The consultants will also be able to refer groups to technical experts or to State officials if their help is needed.

The Office of Planning and Analysis is the logical agency to coordinate the information exchange groups and serve as secretariat. The Management Information Division of the Department of Finance, which has general responsibility for the maintenance of information standards in the state, is the logical agent to exercise technical supervision of data standards for the planning information community.

In many cases, standards set for the benefit of the entire community will result in increased collection and processing costs for the producer. Some readjustment of budget allocations will be necessary if the community is to provide a major channel for planning information. The Bureau of the Budget should therefore be represented on the policy making body.

Once the information community has been established and its information exchange capabilities proved, it will be necessary to require, either by legislation or by executive order, the use of information community channels for information storage, retrieval, and analysis. If use of the information community resources were discretionary, it would be in an agency's interest to use shared information only where such information appears to offer conclusive evidence in favor of its preferred option and against alternative

courses of action. The information community would thereby lose all credibility.

It is not appropriate to suggest an exact structure for an information community in this early stage of research and development. However, the feasibility study indicates that the administrative requirements for establishing and supporting an information community are:

1. a representative body for determining information needs and setting information standards across the entire spectrum of state planning agencies;

2. a mechanism for allocating collection costs incurred in meeting the additional requirements set by registered users;

3. a requirement by the Governor or the Legislature that all planning and administrative information of potential use to more than one agency be channeled through the information community.

CHAPTER 4

THE STATE OF THE ART IN GEOGRAPHIC INFORMATION SYSTEMS

In order to investigate the feasibility of implementing an information system in the State of Illinois, the Institute for Environmental Quality requested that the Center for Advanced Computation survey the state of the art in geographic information systems. The scope of the survey was broad and included systems in use, planned, and under development.

Programs at over 60 universities, governmental agencies, professional study groups, and private companies in six different countries were investigated and are summarized in Appendix B and Appendix C. Over 30 of the groups surveyed had implemented, or were developing, a computer based geographic information system to store and/or analyze area, network, or point data.

Unfortunately, the science of geographic information systems has not developed a good taxonomy. As a result, the term "geographic information system" is used to describe single-purpose graphics programs, geographic data bases, geographic data base description techniques and information systems which can retrieve, reformat, analyze, and display geographically referenced information.

In this document the term "comprehensive geographic information system" will refer to an information system that has the following attributes:

1. The system must possess a data base which can be indexed using a geographic locator.

2. The system must be able to manipulate and analyze the geographic data base by:

   a. displaying raw data,
   b. aggregating raw data into classes,
   c. tabulating the distribution or breakdown of data within a class, and
   d. performing arbitrary arithmetic operations on the data.

3. The system must have the ability to prepare graphic displays which present the results of analyses in map form without writing a special program to prepare the map.

Most computer systems were built as part of a project designed to solve a specific problem and therefore could not be called comprehensive. There was poor communication between the designers and the users of several of the systems. In some cases, competent users underestimated the level of skill required to properly engineer the computer support system. In other cases, competent computer scientists, working without access to users who could adequately guide design efforts, produced elegant computer systems which did not address real world problems.

Few comprehensive geographic information systems are available to problem solving organizations; many have been abandoned due to low user interest or high cost. The cost of system development has varied widely, from $100,000 to over $3,000,000. Many groups paid maximum attention to minimizing the development and operational costs of their systems, a strategy which often required costly new data collection efforts. Investigation revealed a trade-off between system costs and data acquisition costs, with the latter tending to greatly exceed the former. For each dollar spent on the system, it is common to spend another ten dollars on data acquisition. Collecting new natural resource, social, and economic data in the State of Illinois will cost tens of millions of dollars. It is cost-effective to double or even

triple the cost and complexity of the computerized system in order to be able to input and analyze existing data.

Few systems investigated had strong applicability to state problems. Many approaches proved inefficient, too expensive, or inferior to recent developments. More frequently, systems were designed to solve very specific problems, and, as a result, were difficult or impossible to expand. Very few systems addressed the problems of flexibility and general applicability to geographically related problems.

Types of Systems

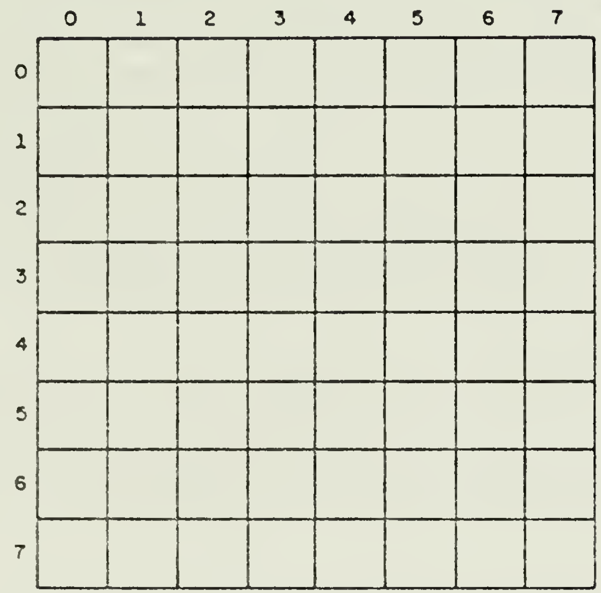
Five types of geographic information systems emerged from the study: uniform grid, parcel, area boundary, network, and point systems. A brief description of the nature, capabilities, and implementations of each type of system follows.

Uniform Grid Systems

Uniform grid systems superimpose a uniform grid, usually based on UTM, latitude-longitude, or arbitrary x,y coordinates, on the study area (fig. 1). A data collection effort is then initiated to determine the attributes of each cell.

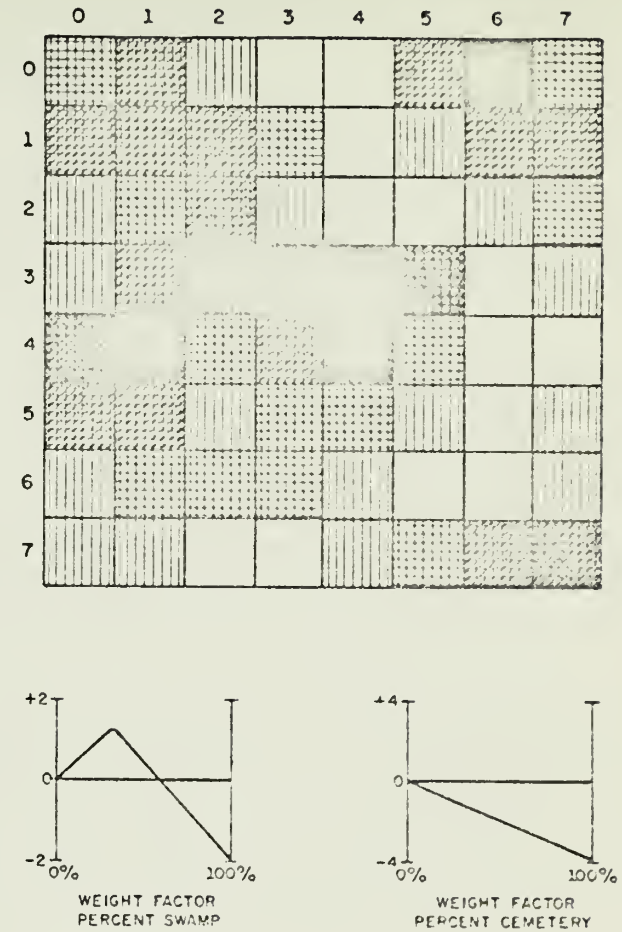
Uniform grid systems are usually capable of overlaying attributes of the grid cells and presenting results in mapped form. Most systems store a single number for each cell attribute -- for example, the suitability for industrialization based on soils, the suitability for industrialization based on local land use, etc., may be rated on a scale from one to ten. The user can then ask to have the suitabilities for industrialization based on soils, local land use, geology, etc., summed for each cell and plotted in map form with increasing density corresponding to greater total sums. Other systems allow a user to sum more general weighting functions of each cell attribute. For example, a weighting function for use in a highway corridor study could be designed to negatively weight a cell if it contained a high percentage of cemetery land or a large percentage of swamp land. (fig. 2).

LUNR [1], LUIS [2], MIADS [3,4], ORRMIS [5], GRIDS [6,7,8], and CMS [9] are examples of uniform grid systems. Of these, LUNR, implemented at Cornell, is the most advanced, using a one kilometer UTM grid to cover the State of New York (140,000 cells). LUNR has a good user language understandable even to untrained programmers. The



A UNIFORM GRID

FIGURE 1



EXAMPLE OF WEIGHTED GRID CELLS
AND WEIGHTING FUNCTIONS

FIGURE 2

language can be used to prepare tabular summaries of data, to analyze data, and prepare weighted, gray scaled maps using the SYMAP [10,11] programs.

Uniform grid systems like LUNR are the easiest systems to design and implement. Storage, retrieval, data overlay, and graphics support associated with these systems are straightforward and can frequently be accomplished with only moderate skill. For these reasons, these systems are usually the first attempted by an organization. If a new data collection effort using aerial photographic interpretation is undertaken, and if only a few data items are to be acquired, the time and costs involved in data acquisition and system development can be significantly reduced by developing a single purpose uniform grid system. This advantage will normally be lost if attempts are made to introduce new capabilities or existing data to the system.

There are drawbacks to using a uniform grid system. Very little data based on any uniform grid system is currently available. Therefore, new data must be obtained. If, at a later date, it were desired to use data on some other grid, perhaps smaller than or slightly displaced from the original grid, a new data collection effort might be required at an additional cost.

Furthermore, uniform grid systems create problems with data since they arbitrarily constrain data within uniform grid lines which do not conform to either natural or political boundaries. Most data used in regional and local planning, conform to politically or naturally defined boundaries such as minor civil divisions, ownership parcels, or watersheds. Due to inaccuracies introduced when aggregating, the uniform grid system cannot accurately represent the fundamental geographic relationships of politically or naturally bounded data.

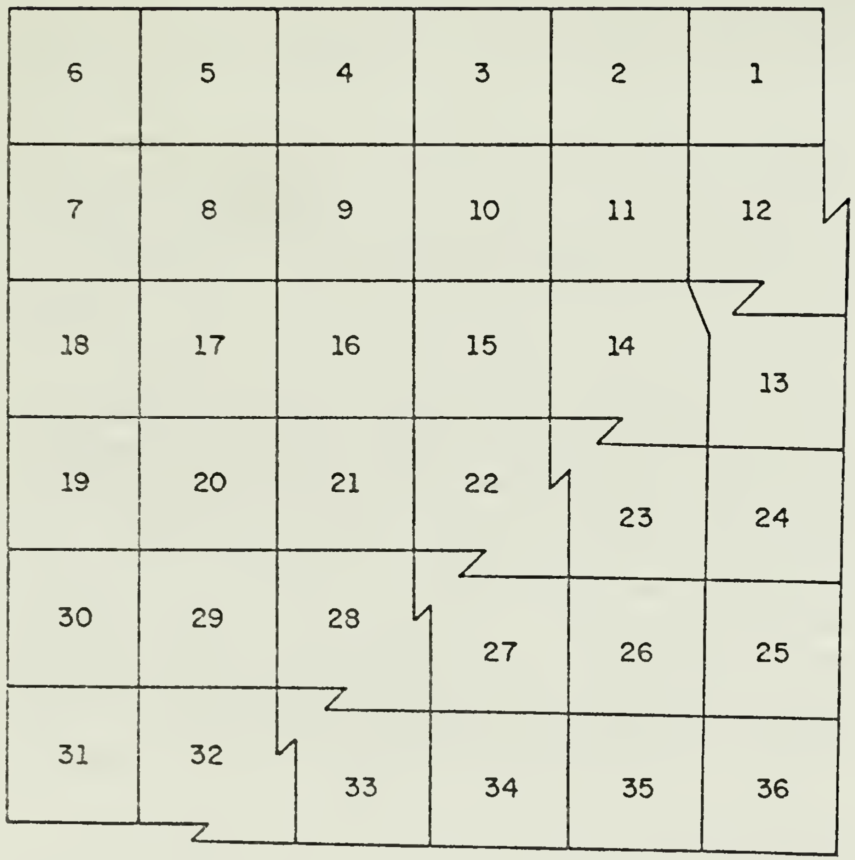
Parcel Systems

Parcel systems permit the collection of data within naturally or politically defined cells, rather than uniform cells. Two parcel referencing schemes are used: non-uniform grid schemes and general parcel schemes. Non-uniform grid systems use a nearly uniform grid to refer to data; they look similar to a uniform grid system but are internally parcel systems.

An example of a non-uniform grid is the rectangular, public land survey which covers most of the nation. The survey system was intended to divide the nation into uniform square mile sections. Due to a large number of survey errors, approximately two percent of the sections are irregularly shaped, have areas other than one square mile, do not join neighboring sections, or have duplicate numbers (fig. 3). Although the vast majority of the sections can be addressed by a regular matrix scheme, additional geolocators must be used to refer to some sections.

General parcel systems accept any system of polygons which cover the study area, and normally assign an arbitrary number to each parcel (fig. 4). Since access to a reference map is required to locate a parcel in such a system, geolocators are often inserted into the data base to allow a study region to be created by data base content. For example, in a system based on census blocks, one might add the ZIP code and mental health catchment code of each block to the data stored for that block.

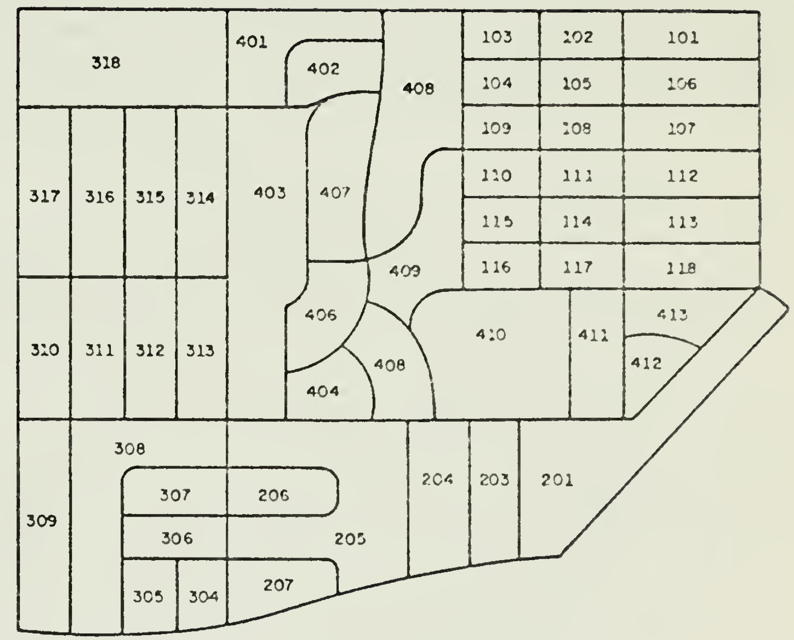
NARIS [12,13], DIME [14,15,16,17,18], MLIS [19], GRDSR [20,21,22,23], GIST [24,25,26,27], FPIS [21,28,29,30,31], and the French Information Network for Regional and Urban Planning 21 are examples of parcel systems. NARIS and MLIS use non-uniform grid referencing schemes based on the rectangular land survey; the others are general parcel



SECTION IRREGULARITIES CAUSED BY SURVEY ERRORS IN
TOWNSHIP 36 NORTH RANGE 13 EAST
(SOUTHWEST CHICAGO) SCALE 1:62500

IN THIS EXAMPLE, EAST MOVING AND WEST MOVING SURVEY TEAMS MET AT AN INDIAN TREATY BOUNDARY AND RESOLVED SURVEY ERRORS ALONG THE DIAGONAL TREATY LINE.

FIGURE 3



SAMPLE PARCEL MAP
(CENSUS BLOCK MAP USED IN A CENSUS DIME FILE)

FIGURE 4

systems. NARIS is representative of and demonstrates most of the capabilities of such systems.

NARIS was developed for the Northeast Illinois Natural Resource Service Center by the Center for Advanced Computation at the University of Illinois as a regional information system designed for use in the eight counties around Chicago. The basic NARIS data collection unit is a quarter-quarter section, 40-acre tract based on the rectangular land survey.

NARIS can produce tabular summaries of the quarter-quarter section data within it and will aggregate data up to quarters, sections, townships or to arbitrarily defined study regions of contiguous or noncontiguous groups of quarter-quarter section tracts. Study regions can be created by describing a specific geographic region or by selecting cells with specific attributes (e.g., a study region consisting of cells having a soil type with no septic limitations). Weighting functions like those used in uniform grid systems are available. Analyses are performed by creating, combining, and intersecting study regions and summing the values of weighting functions applied to the attributes of cells in these study regions. The results of analyses may be graphically displayed as mapped output.

In NARIS, attention has focused on developing an interactive user language. The user language is the result of joint efforts by planners, earth scientists, and computer scientists. It was designed explicitly for the regional planner with no prior computer experience and only brief training. The conversational capabilities of the language allow users to explore and analyze the NARIS data base from remote terminals.

DIME and GRDSR are examples of systems designed to handle census data. GIST includes housing and economic data in addition to census information. Most parcel systems are batch oriented and aimed at

the user who desires to retrieve, aggregate, and tabulate his data rather than weight and map it.

Parcel systems are more difficult to build than uniform grid systems. Although current information retrieval techniques make them no more expensive to operate than uniform grid systems, they remain, unfortunately, more expensive to design and develop. While the regular grid of a uniform grid system makes map preparation trivial, map preparation in a parcel system is more difficult and normally requires that parcel boundaries be digitized.

Many parcel referencing schemes are available, including naturally bounded parcels such as such as river basins and watersheds and politically bounded parcels such as minor civil divisions and various levels of the township and range public land survey system. The public land survey has a special advantage since it provides an easy addressing scheme for any section or quarter-quarter section in the state. It is the legal framework for ownership parcels, and is commonly used by planners and natural resource data collection agencies. The Illinois Geological Survey has already digitized all of the section boundaries in the State for the ILLIMAP system. The ILLIMAP data base can be inexpensively extended to form an excellent basis for a rectangular land survey based parcel system for most states. NARIS uses the ILLIMAP data base to prepare accurate maps of analyzed regions. Large amounts of data already exist for sections and townships, especially natural resource data, which is the most expensive to collect.

Unfortunately, parcel systems share one difficulty with uniform grid systems. If a grid of finer resolution than that on which the data was originally acquired is to be used, new data collection effort conforming to the new parcel grid may be required. But parcel systems have an advantage over uniform grid systems when data collection conforms to natural or political boundaries. In such cases the data

tends to more accurately model the geographic relationships than do uniform grids, making analyses less arbitrary and preserving the fundamental spatial relationships originally recognized when the data were collected.

Area Boundary Systems

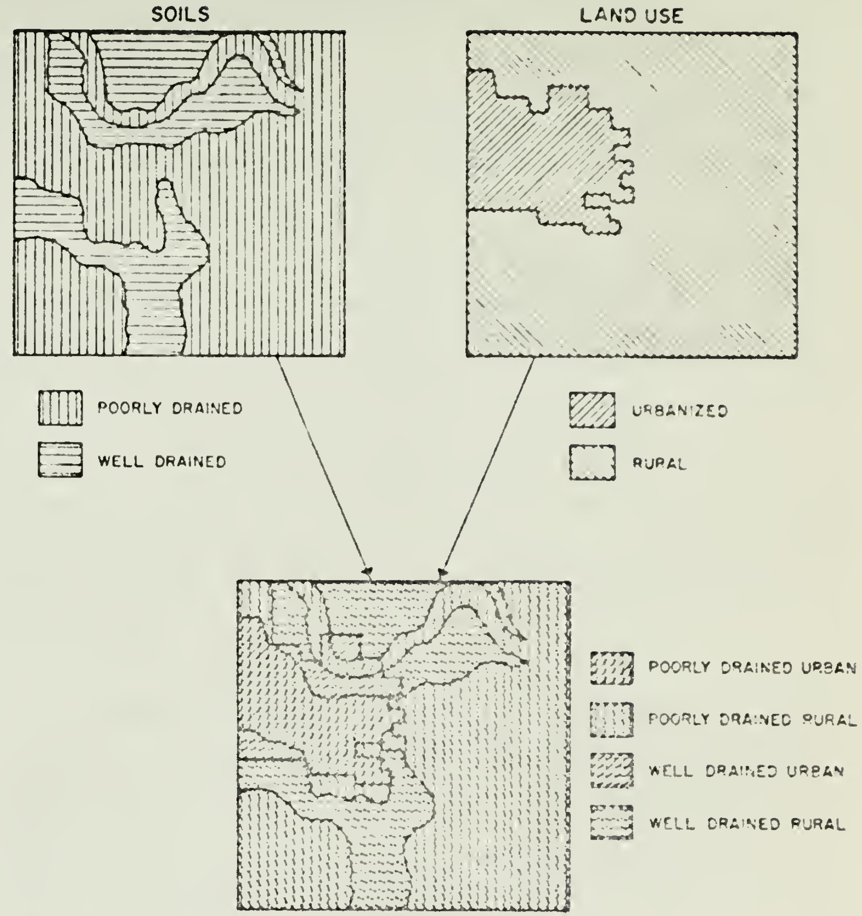
Area boundary systems are the most general and potentially the most accurate representation of area related data. In this system, the boundary of each feature is described using a digitizer or scanner, and the attributes of the area within the boundary are recorded (fig. 5).

Area boundary systems are difficult and expensive to develop and operate. CGIS [26,32,33], a Canadian area boundary system, has cost over $3 million for software development and over $20 million for data collection. The cost of using an area boundary system for any purpose other than simple map overlay preparation is prohibitive.

Like parcel systems, area boundary systems require digitizing efforts in order to prepare mapped output. Unlike parcel systems, it is necessary to digitize area boundaries before data can be input to the data base. For example, it is possible to describe, manipulate, and tabulate the attributes of a minor civil division within a parcel system without providing the digitized data needed to prepare maps, but it is not possible even to input such data to an area boundary system without first describing the geographic boundaries of the minor civil division.

Any geographic referencing scheme is suitable for an area boundary system. Examples include latitude-longitude, state plane coordinates, and arbitrary x,y coordinates.

Area boundary systems have a strong advantage over uniform grid and parcel systems. Virtually all existing data can be processed if one is willing to digitize, or prepare on automated scanners, maps of the



AREA BOUNDARY OVERLAY OF SOILS AND LAND USE DATA
TOWNSHIP 19 NORTH RANGE 9 EAST
(URBANA, ILL.) SCALE 1:150000

IN THIS EXAMPLE, 6 SOIL PARCELS AND 2 LAND USE PARCELS ARE OVERLAYED PRODUCING 21 PARCELS WITH BOTH SOILS AND LAND USE ATTRIBUTES

FIGURE 5

data. Unlike uniform grid or parcel systems, area boundary systems can reference data initially at the finest resolution possible and avoid future data collection efforts.

Upon initial inspection, area boundary systems appear the most capable for data analysis and manipulation, assuming the cost objections could be overcome. However, experience has shown that such systems organize their data too finely. The planner is frequently searching for geographic trends (e.g., Is an area becoming more urbanized?) in his data, which, more often than not, requires some level of aggregation. A planner using an area boundary system may find himself examining trees while looking for forests.

All of the area boundary systems built to date, CGIS in Canada, MAP/MODEL [34,35] in Oregon, and PIOS [36] in California, have input a series of parcel maps for each data type. For example, a soils map describing drainage is input followed by a land use map (fig. 5). Each map can then be overlaid within the computing system. As figure 5 indicates, two relatively simple maps with very few parcels can be overlaid, producing a single map with many more parcels. Note that most of these additional parcels are formed at the boundaries of the original parcels. This is a common experience; land uses tend to follow natural boundaries closely but not exactly. Natural resource boundaries tend to overlay in a similar fashion (e.g. vegetation and soils). The precision of boundary definition determines the extent to which multiple overlays will produce small areas where attributes are incorrect. In tests in a large region in New Guinea, when overlaying polygons of soils, vegetation, and slope, "only 38 percent of the time could they predict that all of the variables said to be present at a single point were actually there" [36, p.47].

The best known and most ambitious area boundary system, CGIS, intended but failed to implement a point or network capability. The CGIS system provides little more than an area boundary base file and a

few subroutines for performing rudimentary overlays of those base files. A PL/1 program is necessary to retrieve or analyze data, requiring a trained programmer who can interpret requests and write a program to retrieve data from CGIS tape files.

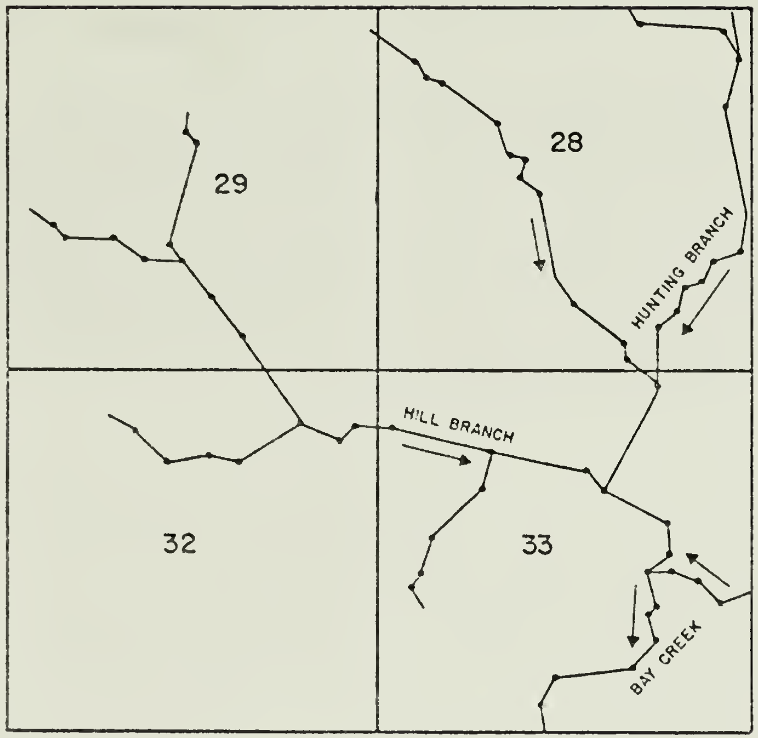
CGIS is exorbitantly expensive and provides little analysis capability. The computer system suffers from two basic faults: the difficultyinvolved in attempting to manipulate and analyze area boundary data, and a poor software system design. If the area boundary files were used as a raw data base to aggregate into a uniform grid or parcel system for manipulation and analysis, operational costs could be drastically reduced. Other area boundary systems have been constructed, most notably MAP/MODEL and PIOS (based on MAP/MODEL) which, while they still suffer from the fundamental difficulties of analyzing area boundary data, use much better and more efficient access algorithms.

Network Systems

Network systems store the attributes and geographic location of network links and nodes (fig. 6). The U.S. Environmental Protection Agency STORET [37,38,39,40] project is one of the more interesting network projects. STORET will have great impact on future network systems, since all of the stream networks in the country will be digitized for the program. This will create a rich network data base to support water quality planning programs.

Preliminary estimates indicate that the cost of the developing and operating a computer-based network information system is similar to the cost of developing and operating a parcel information system.

Network systems are similar to parcel systems in that they require digitizers or scanners to prepare mapped output, but they are still capable of manipulating network data and preparing tabular summaries of network data without digitizer input.



STREAM NETWORK DATA
SECTIONS 28, 29, 32 & 33 OF
TOWNSHIP 11 SOUTH RANGE 5 EAST
(SHAWNEE NATIONAL FOREST) SCALE 1:24000

FIGURE 6

Any geographic referencing scheme is suitable for describing networks. Since networks are described by the point locations of the terminal nodes of each link, the problem of geographic referencing is reduced to the problem of point referencing. Standard mathematical expressions describe the relationship of latitude and longitude, Universal Transverse Mercator, Lambert, state plane and other coordinate systems. In addition, the existence of the ILLIMAP data base for Illinois allows the conversion of most legal descriptions of points (e.g., well locations) to Lambert coordinates and hence to any other point coordinate system. Since coordinate conversions are trivial, the user need not be arbitrarily constrained to use any one convention.

A network system in Illinois would be used primarily to manipulate stream and transportation network data. In particular, the wealth of data in the Illinois Department of Transportation roadway file could be manipulated in a network system if the end points of links in the roadway network were digitized. In addition, the U.S. Environmental Protection Agency is digitizing the stream network in Illinois to prepare STORET base files which can then be used in problem oriented research and program development in water quality management.

Point Systems

Point systems store data only as attributes of specified points. Like uniform grid systems, point systems are relatively easy and inexpensive to develop and operate. They are frequently developed to avoid building a parcel system, by listing area-related data as the attributes of parcel controids. As a result, some systems (e.g., GIST in New York) could be classified as point systems if any point specific (as opposed to parcel related) data were stored. Some tabular summary capability is available in point systems, but graphic output is limited to the generation of dot maps.

Unlike a uniform grid system, a point system does require digitization of the parcel centroids or the actual point location for which data is stored in order to accurately prepare mapped output. As indicated in the description of network systems, any geographic referencing scheme is suitable for describing point data.

Point sources of air and water pollutant emissions under permit from the State Environmental Protection Agency are located by latitude-longitude coordinates. Locations of oil, gas, and water wells, and related data are examples of point data, as well as most meteorological observations and measurements of air and water quality. Point system capabilities are required to support air quality plans and provide data for air dispersion models.

The Suitability of Specific Geographic Information Systems for Illinois

Descriptions of all the geographic information systems investigated are contained in Appendix A. We have isolated several systems in this section which have powerful capabilities which should be explicitly brought to the reader's attention and/or which are popular systems that are frequently mentioned in the literature and may thus raise some questions in the reader's mind as to their applicability in Illinois.

LUNR (Land Use and Natural Resources)

The LUNR system was developed for the Office of Planning and Coordination in the State of New York by the Cornell Center for Aerial Photographic Studies. LUNR is clearly the best of the uniform grid systems analyzed. In particular, it has all of the capabilities and notable extensions beyond the capabilities of the GRIDS system developed in southern California for manipulating census data, the GRIDS system developed in Washington to handle natural resource data, the MIADS system developed to handle forestry data, and the Composite Mapping System developed by the Economic Development Administration of the Department of Commerce.

The LUNR system cost over $750,000 to develop. Most of that cost was involved in data acquisition. Approximately $100,000 was involved in computer system generation. The LUNR system contains land use data prepared from aerial photographic interpretation on one square kilometer UTM grids. It has a good user language which can be used by an untrained programmer to prepare tabular summaries of data, analyze data, and prepare weighted, gray scaled maps using the SYMAP programs.

Because LUNR is a uniform grid system, it suffers from the inherent disadvantages of such systems--notably data collection costs

and data representation. The current LUNR system does not handle point or network data and no plans exist for implementing point or network capabilities. If a uniform grid system were acceptable for the State of Illinois, we would recommend that the LUNR system be acquired and used for these purposes since it is the best available system.

The only significant alterations that we would recommend to the existing code would be in the area of data compression and data access techniques. The current system can store only multiple attributes of each square kilometer cell. Space is allocated for all possible attributes of each cell. For example, space is reserved for information about unique sites, the linear footage of streams, and the number of acres of lake in each cell even if none of these exist. In fact, only a small percentage of all possible attributes ever apply to a single cell. A simple data compression scheme that only records data that exists and does not leave space for data that does not apply to a cell is needed. In addition, an augmentation to record tree structured data for each tract would be a powerful addition to the LUNR system.

NARIS (Natural Resource Information System)

NARIS was developed for the Northeast Illinois Natural Resource Service Center by the Center for Advanced Computation at the University of Illinois. NARIS is a parcel system that uses as its fundamental data collection unit a quarter-quarter section, 40-acre tract based on the rectangular land survey. NARIS is a regional information system designed for use in the eight counties around Chicago.

The NARIS computer system was developed at an approximate cost of $200,000. Funding was provided by a Ford Foundation grant, supplemented by the Northeastern Illinois Planning Commission and the University of Illinois.

Since NARIS is a parcel system it was more expensive to develop than a uniform grid system. However, it was amenable to using data already collected or being collected by existing agencies. This significantly reduced the data collection costs. Only $26,000 were originally supplied by the Ford Foundation for data collection. All of this money has been used to influence agencies to collect data in a NARIS format and to provide seed money for data collection activities. NARIS does not have any point or network capabilities. It does, however, pay very close attention to the large size of its data base and employs state of the art information retrieval techniques to compress and encode this data for rapid retrieval and minimum storage requirements.

NARIS can produce tabular summaries of the quarter-quarter section data within it. This data may be aggregated up to quarters, sections, townships, or arbitrarily defined study regions. NARIS study regions consist of contiguous or noncontiguous groups of quarter-quarter section tracts. The results of analyses performed by the NARIS system may be graphically displayed as mapped output. NARIS uses the ILLIMAP data base to prepare accurate maps of the eight counties.

Great attention has been paid in NARIS to the development of an interactive user language. The NARIS user language is the result of joint efforts by computer scientists, earth scientists, and planners. It was designed explicitly for the regional planner. Its use requires no prior computer experience and only a short training period. The conversational capabilities of the language allow a user to access the NARIS data base in exploratory fashion from a remotely located terminal. A documented, first version of NARIS will be operational in July, 1972.

NARIS is an acceptable interim geographic information system that can be used by metro-regional planners in the Chicago area. In its present form, it is not suitable for handling data other than quarter-quarter section parcel data.

CGIS (Canadian Geographic Information System)

The Canadian Geographic Information System is a very large scale area boundary system developed by IBM for the Canadian Department of Regional Economic Expansion. The Canadian Geographic Information System is a serious failure. Over $3,000,000 in the computer system and over $20,000,000 in data acquisition have already been invested.

CGIS intended to, but never did, implement a point or network capability. CGIS has no user interface language. Basically, the CGIS system provides little more than an area boundary base file and a few subroutines to perform overlays of those base files. In order to retrieve or analyze data the user must write a PL/1 program. This requires that he have a trained programmer at this disposal who can interpret his request and write a computer program to retrieve the data from the CGIS tape files.

Staff interviewed at the Canadian Land Inventory are very disappointed with the system. They feel that CGIS is much too expensive to be used by the provinces.

The computer system suffers from two basic faults. First, it attempted to manipulate and analyze raw area boundary data rather than aggregate the raw data base into a uniform grid or parcel system for manipulation and analysis. Second, the quality of the computer science work was poor.

Data input costs were clearly out of line. The specially designed scanner hardware and software to input and correct maps was approximately 10 times more expensive than any other map input technique investigated in the state of the art survey (approximately $500 versus a more typical $40 to $50 per map).

In addition, the basic software system design, especially the choice of retrieval algorithms and area boundary search algorithms, was poor. The overlay algorithms chosen use large amounts of core and central processor time. Other area boundary systems have been constructed, most notably the MAP/MODEL system in Oregon. While MAP/MODEL still suffers from the fundamental difficulties of analyzing area boundary data, it does use much better and more efficient access algorithms.

MIDAS (Maine Information Display and Analysis System)

The MIDAS System was built for the Maine Bureau of Inland Fisheries and Game by NARISCO. The work was originally started by CAI, Computer Applications, Inc., which was later purchased by NARISCO. MIDAS is a general purpose tape data management system which contains no general mapping or analysis capability. MIDAS is capable of retrieving data and preparing tabular summaries. The system also has an ability to prepare special purpose files for external statistical analysis using the BIO-MED statistical system.

A geographic data manipulation capability is provided in the system by appending a 200 byte, master block index to each data record stored. The addition of this index approximately doubles the size of the average record. Contained within this 200 byte index are the minor civil division code, county code, UTM coordinates, and other geographic locators which apply to that particular piece of data. MIDAS is not an area, point, or network geographic data manipulation system. It is better classified as a tape data management system with geographic information appended to its data base.

The MIDAS software will cost in excess of $300,000. MIDAS is a good system with a well designed user interface language. The State of Maine appears pleased with the system and it seems adequate for Maine needs. While the system was competently designed and developed, the

State of Maine probably could have saved the major portion of the $300,000 development costs by purchasing an "off the shelf" tape data management system and augmenting it, where necessary, to perform specific tasks.

The MIDAS system is inadequate for use in the State of Illinois as a comprehensive geographic information system. In particular, it has no mapping capability, it has no geographic analysis capability, and it is coded in assembly language for RCA computers. In practice, the smallest geographic unit of resolution is the minor civil division. Although point data is stored, it is typically accurate only to the minor civil division; the centroid of the minor civil division is used as the point locator. Since RCA is no longer in the computer business, the State would be ill advised to acquire an assembly language coded system for an RCA computer.

SYMAP

SYMAP is a computer graphics program developed by the Harvard University Graduate School of Design, Laboratory for Computer Graphics and Spatial Analysis. SYMAP is often confused with a comprehensive geographic information system. As indicated in the SYMAP documentation, "SYMAP is a computer program designed ...to produce low cost graphic displays as spatial patterns using standard computer line printers." SYMAP is only a graphics package, not an information retrieval system or information management system.

The SYMAP program is written in FORTRAN and is designed to be exported to any modern computer which has a FORTRAN IV compiler. While SYMAP was originally designed to produce inexpensive maps on readily available line printers, a new version, SYMVU, is available which can also produce better quality maps on pen plotters and CRT plotters.

For the 1970 census, the Bureau of the Census undertook a study of mapping programs which involved their use under actual working conditions. Census Use Study Report #2 on computer mapping favorably reviewed the SYMAP program. The following quote from that report is the only major criticism raised.

> Within the SYMAP System changing from one type of map, using one mapping option, to another type of map is not easy. For instance, switching from a map using irregular polygon areas, as in conformal block shading, to a map using single data points as in contour shading requires the input of a completely rearranged geographic base. In addition, the visual attractiveness of maps produced by SYMAP was sometimes disappointing and local users tried to add cosmetics by hand to make them more presentable as well as facilitate orientation.

Note that this criticizes the limitations of line printer graphics and attempts to use graphic programs, without the support of comprehensive data management facilities, rather than SYMAP itself.

SYMAP is frequently used as the graphics package for many geographic information systems. Most uniform grid systems and some parcel systems do not prepare their own graphic output. They prepare command streams and data files to be processed by the SYMAP program. SYMAP has had significant programmer time invested in it, produces good quality maps, and implements many desirable graphics facilities on any machine that has a FORTRAN compiler and a line printer. The purchase price of the SYMAP programs is nominal ($484.00). SYMAP should be used to prepare line printer graphics in the IRIS system.

DIME Files (Dual Independent Map Encoding)

The DIME system is also frequently believed to be a geographic information system. It is not a computer system, nor is it a computer program. The Census Bureau provides no generalized computer software to support it. The DIME system is a technique for encoding and storing census data.

The DIME system records the end points of census block faces and is used to describe census blocks and tracts. The end points of a DIME file are coded using arbitrary coordinates. It is not necessary to tie these x,y coordiantes to any existing earth related coordinate system. Latitude-longitude, state plane coordinates, and inches and feet measures from the map used by the data encoder have all been employed in generating DIME files.

While the Census Bureau does not supply generalized computer software, it does have UNIVAC 1108 software to handle its own DIME files internally. In addition, many private consultants and governmental organizations have prepared their own software for manipulating DIME files.

During the New Haven Census Use Study, several computer programs were developed to support the DIME concepts. These programs were written for an IBM/360 computer and vary widely in reliability and usefulness.

The most commonly used DIME "programs" are ADMATCH and the Address Coding Guide. The address coding guide is a data file that describes the street name and the beginning and ending addresses of each block face in a census block. Theoretically, the ADMATCH computer program will match street addresses with census blocks using the Address Coding Guide. Since people frequently use non-standard abbreviations and misspell street names, the preparation of an ADMATCH program is actually far more troublesome than would first be indicated. Some agencies claim they have been able to write or use an ADMATCH program,

but the majority have reported great difficulty in getting the program to function properly.

Whatever system is accepted by the State of Illinois, it must have the capability of manipulating DIME files.

Others

There are several other systems that the reader might hear mentioned that are not covered in detail in this section. In particular, the reader is referred to the descriptions in Appendix B of the ILLIMAP system. ILLIMAP is a base file and a mapping program for preparing section maps from township, range, and section descriptions in the State of Illinois. IIPS (Inter-Institutional Policy Simulator) at the University of British Columbia in Vancouver is a Ford Foundation project which is attempting to integrate five simulation packages around a powerful, graphical display system. The bulk of the IIPS information processing capabilities are still in design. The Minnesota Land Use Information System (MLIS) is a forty acre quarterquarter section system which is still in the preliminary design phase and may be similar to the NARIS system. The STORET System has a digitized stream network base file. This system, supported by the federal EPA, can prepare tabular summaries of the network data for use in monitoring and enforcement programs.

Where Do We Go From Here

Problems to Avoid

Over thirty organizations have already invested millions of dollars in computer systems for analyzing geographic information. Most of these systems have failed to become a commonly used decision making tools. In each case, the failure is normally due to all of the following reasons:

First, the systems were too small in terms of capabilities and data. The major efforts have normally cost $100,000 to $300,000 for software and several times that for data. This investment created a tool of use in a specific problem area. The tool could not be readily enhanced to provide new capabilities and handle new data without incurring significant costs. Therefore, when user subsidies disappeared, problem priorities shifted, new data was needed, or bureaucracies were reorganized, the system was abandoned. It did not have the broad capabilities to support the diverse clientle which would insure its continued economic viability. A good example is the New York LUNR System. Although LUNR is the best of the uniform grid systems, its analysis capabilities and land use data base were inadequate to survive the reorganization of its sponsoring state agency.

Second, the state-of-the-art is very poor. The language concepts, data management techniques, graphics facilities, and analysis packages employed in those 30 systems were usually very old computer technology - at a time when innovation in all these areas of computer technology was essential to produce a viable tool. The computer science skills required to build a good system were often underrated by the bureaucrats and planners who sponsored the systems. Many of these systems were designed and implemented by students or civil servants with only one or two years of experience. As a result they used naive and inadequate techniques to attack difficult technical problems and were

not capable of generating new technology when needed. The results have been catasrophic in some cases. A case in point is the Canadian Geographic Information System (CGIS). CGIS was originally to cost $400,000 (½ for special scanner hardware and ½ for software). Due to a poor ability of the contractor to produce the needed technology, over $3,000,000 has been invested to date by Canada and CGIS still has no user language, has only rudimentary data manipulation capabilities, and is so exhorbitantly expensive to operate that it is virtually useless. In addition, over $20 million has already been invested to collect data specially formatted for this system.

Third, the integration with government was inadequate. A total commitment is required on the part of the state, regional, or local government involved. In addition, the client group must be carefully studied to determine what innovations in technology and organizational structure are socially and politically acceptable in the client group and in what order, over what time scale these innovations should be introduced. The failure of the Integrated Municipal Information System (IMIS) in Charlotte North Carolina is an example. Most of the resources of the IMIS project were spent on the computer system in the hope that somehow online CRT's and terminals would make city government more efficient. Unfortunately, no social or procedural environment existed that was capable of supporting these innovations and no programs were undertaken to create an acceptable environment.

The analysis of existing systems clearly indicates that a major new effort is needed. Another small one or two hundred thousand dollar program can not be expected to achieve the critical mass needed to make it self supporting. That critical mass requires a system with powerful analysis capabilities, useful in many existing problem areas, capable of expansion into new areas, and loaded with an initial data base broad enough to be useful to many clients.

Needed Advances in the State of the Art

There are six advances needed in the State of the Art. These are included in the design specification for the Illinois Resource Information System.

Large remote user community: To date, geographic information systems have been used in a batch environment or by a very small number of conversationally attached users at rather expensive remote terminals. NARIS is the first designed to be handled by a moderate number of remote users via inexpensive Teletype-like terminals. The next advancement should support dozens of conversational users having exploratory access to the geographic data bases via inexpensive terminals used much as desk calculators or slide rules are now used by engineers.

Points, networks, and areas: Many systems have been built which represent either network, point, or area data, but no single system can yet manipulate all three types. The next technological advance should support a system which can handle all three types of data and provide overlay of one type on another. For example, the user must be able to overlay land use data from an area file onto a stream network file in order to determine non-point effluent sources of water pollutants.

Multiple data bases: Technology can be extended to support multiple data bases at many levels of geographic resolution. Separate network data bases, a separate county data base, a separate quarter section data base, and a separate river basin data base should all be stored, retrieved, and analyzed by a single system. Indexing schemes must be designed to enable the user to extract data from one data base at one level of geographic resolution and reaggregate it so it can be compared with data from a data base at another resolution. For instance, one will be able to extract data from a quarter-quarter section data base and aggregate it up to a watershed level for use with other area-related data for that watershed.

Multiple analysis packages and user languages: In the past, geographic information systems have served only one major user or one major activity. Since several classes of users are interested in manipulating the same geographic data, different user languages must be developed. The jargon, the analyses, and the very train of thought used by a highway planner is different from that used by a social planner. It is not possible to fill both needs with a single user language which will still make the computer relatively invisible to the problem solver. The user could not care less about a common data manipulation language.

A major part of the proposed research program will be focused upon the construction and coordination of separate interface packages to perform problem specific analyses for each class of users on the system. Each package will have an interface language based on terminology normally used by the users in that area. Familiar terminology and familiar phrasing will significantly reduce training time and initial reluctance of working planners.

Multiple data entry facilities: Most geographic information systems have in the past been provided with a single data entry mode. For example, data have been input only via punched cards or digitized maps prepared from a single series of aerial photographs. It is feasible to provide a multi-input data entry facility that can accept data from card files, digitizers, scanners, etc., and perform the necessary file reformatting to make those files acceptable to the geographic data base.

Graphics: Geographic information systems have in the past relied on crude line printer graphics or pen plotter graphics. Line printer graphics require felt pen cosmetics to be useful to the planner [14], and the production of variable density maps with pen plotters is tedious and expensive.

NARIS now uses an electrostatic plotter to produce pen plotter resolution maps with line printer ease, speed, and economy, and IIPS [41,42,43] is experimenting with storage tube terminals to provide interactive graphics.

Color provides an extra, needed dimension that is not available in black and white maps. The results of spatial analyses are complex and difficult to present to professional or lay audiences.

Summary

Geographic information systems store and analyze point, network, and area data. No system can yet manipulate all three types of data. Current planning activities require that point, network, and area data be available to the user for analysis within a single system. Parcel systems are preferable to uniform grid or area boundary systems for the normal storage and analysis of area data, but area boundary systems are best for geographic data archieves.

Geographic information systems are being actively investigated by many countries, universities, professional groups, and state and local governments. Most projects have been small single purpose efforts that have failed to be integrated into the working fabric of government. They have failed because they were too small. They could not support the broad clientele which would insure their continued utility in an environment of changing priorities and agency reorganization. They did not take advantage of and build on advanced computer technology. Finally, they were not completely integrated with government from the start.

The most profitable and needed advances in the geographic information system art should be the development of capabilities to support:

1. a large remote user community;
2. point, network, and area data within a single system;
3. multiple data bases representing multiple levels of geographic resolution;
4. multiple problem specific interfaces to the data base each with its own user language;

5. multiple data entry facilities; and

6. advanced graphics capabilities.

References

1. Belcher, D., et. al. New York State Land Use Natural Resources Inventory Final Report, Vol. 1-5, Center for Aerial Photographic Studies, Cornell University, Ithaca, New York, 1971.

2. MacConnel, W. L., and Howard, R. A., "LUIS-Map Information Storage Manipulation, Retrieval, and Display System," ACSM March 7, 1971, p. 349.

3. Admison, E. L., "A Computer Oriented System for Assembling and Displaying Land Management Information." U.S. Forest Service Research Paper PSW 17, 1964.

4. Store, T.G., et. al., "INFORMAP--A Computerized Information System for Fire Planning and Fire Control." USDA Forest Service Research Paper PSW 17, 1969.

5. The Environment and Technology Assessment, Progress Reports, December, 1970, Oak Ridge National Laboratory, February 1971.

6. Harding, R. A., "Resource Inventory by GRIDS." Technical Services Division, Department of Natural Resources, Box 160, Olympia, Washington, 98501, April, 1969.

7. Harding, R. A., DNR Technical Services Manual, 9102 05 GRIDS--A Manual for Data Entry, Department of Natural Resources, Box 160, Olympia, Washington, 98501, April, 1969.

8. "GRIDS--File Building and Editing," internal document, Department of Natural Resources, Box 160, Olympia, Washington, 98501, April, 1969.

9. Nef., G., "A Composite Mapping Ssytem for Practical Location Research," Economic Development Adminstration, U.S. Department of Commerce.

10. "Program Description and Availability Memorandum for SYMAP", Graduate School of Design, Harvard University, November 1971.

11. "A Summary of SYMAP V Subroutines", Graduate School of Design, Harvard University, December, 1971.

12. "NARIS: A Natural Resource Information System." Center for Advanced Computation, University of Illinois, Urbana, 1971.

13. McTeer, Wm., et. al., "NARIS User's Manual", internal documentation, Center for Advanced Computation, University of Illinois, February, 1971.

14. "Census Use Study Report #2, Computer Mapping". U.S. Department of Commerce, Bureau of the Census.

15. "Census Use Study Report #4, The DIME Geo-coding System." U.S. Department of Commerce, Bureau of the Census.

16. "Census Use Study, Ad-match Users Manual." U.S. Department of Commerce, Bureau of the Census.

17. "Geographic Supervisors Manual." U.S. Department of Commerce, Bureau of the Census. Geo 70-108, May 1969.

18. "Geographic DIME File Coders Manual." U.S. Department of Commerce, Bureau of the Census. Geo 70-108, May, 1969.

19. Personal Discussion with G. Orning, "Minnesota Land Information System," summer, 1971.

20. "Geocoding Facts by Small Areas." Bulletin #1. Dominion Bureau of Statistics, Ottawa 3, Ontario. February, 1969. 21. Tomlinson, R. F., Editor, "Environmental Information Systems", Ontario, September, 1970.

21. Tomlinson, R.F., Editor, "Environmental Information Systems" Ontario, September 1970.

22. "Statistic Canada--Introduction to the GRDSR System", Presented at the Provincial Census Data Access Workshop, Ottawa, October 17, 1971.

23. Ion, R. J., et. al., "GRDSR Geographical Referenced Data Storage and Retrieval System", Dominion Bureau of Statistics, Ottawa, 1970.

24. Amsterdam, Robert M., "GIST, A Geographic Information System for New York City (preliminary design)." Office of the Mayor, New York City, July, 1968.

25. Amsterdam, Robert M., "Development of New York City's Geographic Data Network". Spring Joint Computer Conference, 1969.

26. "GIST, New York City's Geographic Information System-Initial Operating capabilities and Future Development". Office of the Mayor, New York City, September, 1970.

27. Amsterdam, R., An Introduction to GIST, New York City's Geographic Information System, Office of the Mayor, May, 1971.

28. Alfredsson, B., et. al., "A Spatial Information System Introduction." FRIS C:1, Sundbyberg, Sweden, October, 1970.

29. Selander, K., "A Spatial Information System, Registration and Storing of Coordinates", FRIS C:1, Syndbyberg, Sweden, October, 1970.

30. Olsson, A. and Selander, K., "A Special Information System Dot Map by Computer". FRIS C:2, Syndbyberg, Sweden, May, 1971.

31. Olsson, A., "A Spatial Information System Program for Coordinate Processing". FRIS C:3, Sundbyberg, Sweden, May, 1971.

32. McClellan, J. B., "The Land Use Sector of the Canada Land Inventory", Geographic Bulletin. 1965. Vol. 7, No. 2, pp. 73-78.

33. Tomlinson, R. F., "An Introduction to the Geo-Information System of the Canada Land Inventory". Canada Department of Forestry and Rural Development. Ottawa, 1967.

34. Arms, S., "MAP/MODEL System-Technical Concepts and Program Description". Columbia Region Association of Governments, Portland, Oregon.

35. Arms, S., "MAP/MODEL System--System Description and Users Guide". Bureau of Governmental Research and Service, University of Oregon, May, 1970.

36. Taylor, L., "Polygon Information Overlay System", Comprehensive Planning Organization, San Diego, California, September 15, 1971.

37. Green, R. S., et. al., "Data Handling Systems and Water Pollution Control". Sanitary Engineering Division Proceedings

of the American Society of Civil Engineers, February, 1966. p. 55.

38. "Location coding for the STORET System" (Revised November, 1968). U.S. Department of the Interior, Federal Water Pollution Control Administration, Division of Technical Support, Pollution Surveillance.

39. "AUTOMAP--A System that Determines River Mileages and Latitude/Longitude for Input to the EPA/OWP's STORET System for Retrieval of STORET Information in Hydrological Order", Environmental Protection Agency, Office of Water Programs, Washington, D. C.

40. Taylor, P., "STORET--A Data Handling System in Water Pollution Control". ASCE Annual Meeting, October 13-17, 1969, Sanitary Engineering Division.

41. Parker, J. L., "Graphics and Information Retrieval Supervisor for Simulators". Department of Computer Science, University of British Columbia, July 28, 1971.

42. Parker, J. L., "The Natural Information Systems Project, An Overview". Department of Computer Science, University of British Columbia. August 8, 1971.

43. Parker, J. L., "Information Retrieval and Large Scale Geographic Data Bases". Department of Computer Science, University of British Columbia. Report #1, June, 1971.

CHAPTER 5

THE PROPOSED IRIS COMPUTER SYSTEM

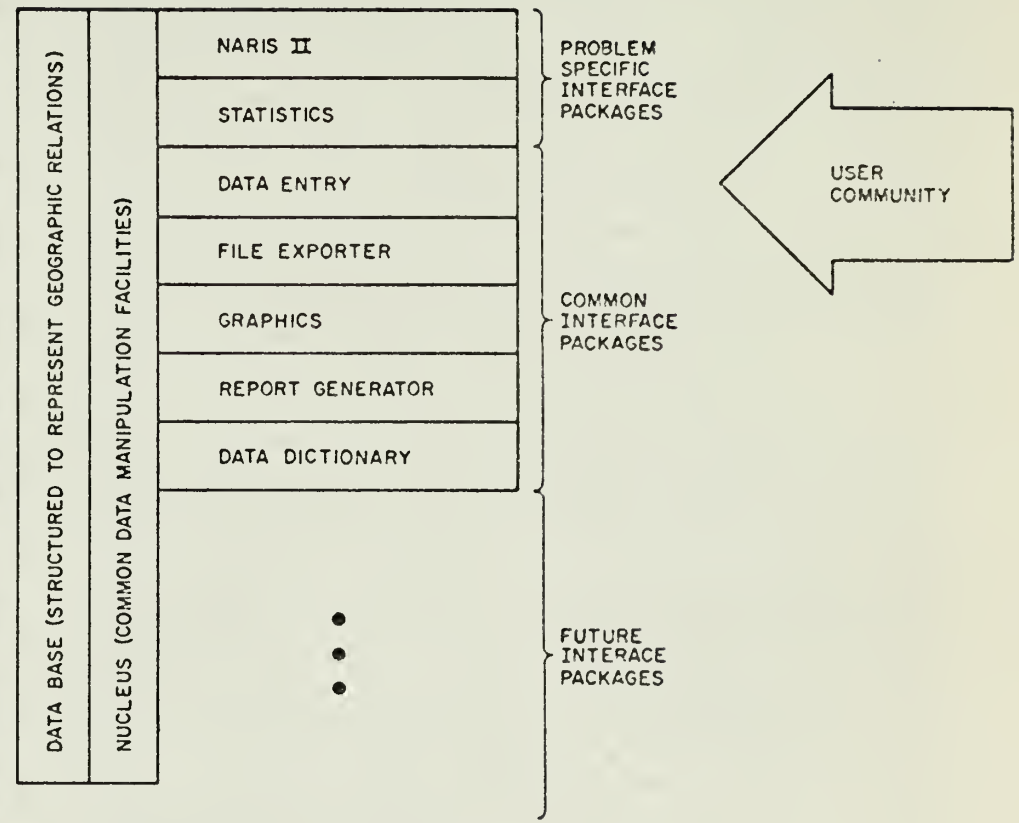
The proposed IRIS system will store data on counties, incorporated townships, survey townships, sections, quarter sections, quarter-quarter sections, ZIP code regions, river basins, minor basins, sub-basins, minor civil divisions, congressional and legislative districts and census units (tracts, blocks, and county subdivisions). Point and network data will also be stored. Any social, economic, natural resource, mechanical, or other attributes of these areas, points, and network links could be stored.

The IRIS system will retrieve and display, in tabular, graph, or mapped form, data relevant to points, networks, and areas. IRIS will be capable of transforming and scaling its data base into a point, network, or area oriented file for input to existing modeling and simulation facilities. It will be possible, in fact fundamental to the IRIS design philosophy, to add analytical packages as required.

The proposed IRIS design recognizes that the system:

1. must be conversationally oriented;
2. must provide a method for preparing files from the IRIS data base that are compatible with other computer systems;
3. cannot provide all analytical capabilities for all users within one user language;
4. must take into account that some user requirements are in conflict with others and therefore require separate interfaces to the data base.

Figure 1 shows the initial software configuration of the system. The Nucleus provides access to point, area, and network data in the data base, and it provides those low level system facilities



INITIAL IRIS SOFTWARE CONFIGURATION

FIGURE 1

required by all interface packages. The message switching facility which allows interface packages and the users to communicate, the facility to create information keys, and facilities to map point, network, or area data into other modes are all part of the Nucleus.

The IRIS data base cannot be just a collection of data files inserted into a general purpose data management system. The data base must be carefully structured to represent fundamental geographic relationships between point, line, and area data. Furthermore, it must be optimized for high-speed retrieval and analysis of large amounts of data rather than for data update and insertion activities.

The interface to the Nucleus and the data base is through common and problem specific interface packages, each having its own language. Thus, each problem specific interface package may have a different user language which mirrors the jargon and analyses characteristic of the problem. The various languages can be as parallel and as similar as is reasonable, but the general intent is to make the single package user comfortable. The multiple package user will tend to be more sophisticated and therefore better able to work in more than one language.

After eighteen months, a first version of IRIS will be released to users for experimentation. Two problem specific interface packages are scheduled for the first release: NARIS II and a Statistics package. The NARIS II package will include the capabilities of NARIS plus an extended study region manipulation capability, an ability to handle quarter section, section, river basin and other data bases in addition to the quarter-quarter section data base of NARIS [2,3]. The capability to manipulate line and network data and to map any of the three data modes (point, network, and area) into any of the other modes will also be provided. The Statistics package will provide standard statistical analyses for operation on files prepared directly from the IRIS base by the Statistics interface or on files prepared by other interface packages (e.g., NARIS II).

Five common interface packages will be provided for the first release: Data Entry, File EXporter, Graphics, Report Generator, and Data Dictionary. The Data Entry interface is composed of many subprograms which enter data into the IRIS data base, correct errors and update the data base. The File Exporter prepared files to be run on software packages external to the IRIS system. The Graphics interface package provides general graphics support for the NARIS II, Statistics, and any other common or problem specific interface package. The Report Generator provides general report formatting and preparation support for the other interface packages. The Data Dictionary interface package indexes the data bases and describes the facilities available within IRIS. Users will normally access the common interface packages indirectly via problem specific interface packages.

In summary, the first release IRIS will be able to store and retrieve social, economic, and natural resource data relevant to points, networks, and areas on a variety of geographic regions, both naturally and politically defined. The system will be able to manipulate data, weight it, sum it, aggregate it, etc., and present the results in the form of tables, graphs, and maps. The initial system will support users who must simultaneously manipulate social, economic, and natural resource data (e.g., planners at the state environmental agencies and the regional planning commissions). The first release will provide raw data and custom formatted and organized files from the IRIS data base for use on other software and hardware systems. Finally, it will be designed to permit the addition of modeling, simulation, and other packages needed by the user.

Design Considerations

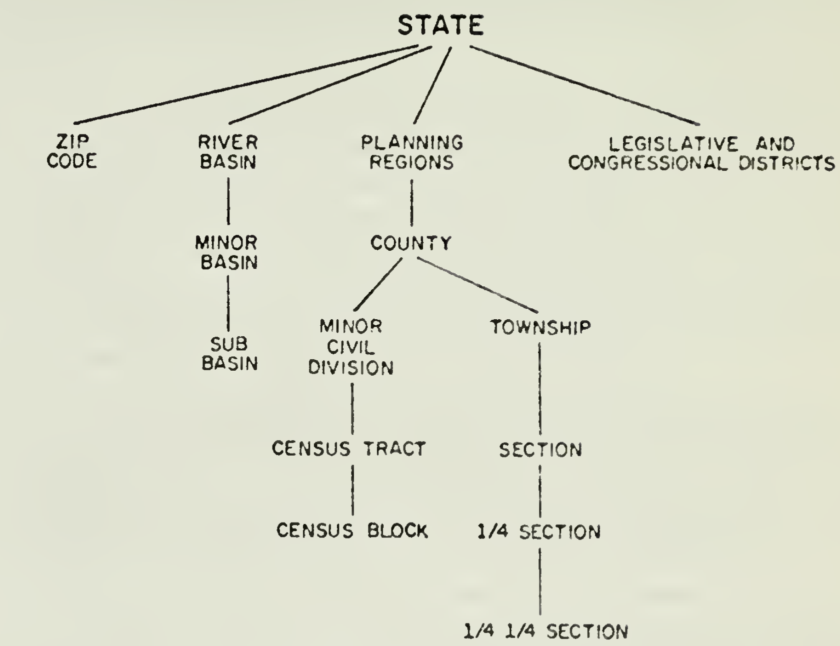
The Data Base

IRIS must provide a capability for storing and retrieving attributes of points, networks, and areas. Furthermore, the data base must represent the fundamental relationships between data that apply to a network, data that apply to a point, and data that apply to an area.

Under user command, IRIS must be able to map one form of data onto one of the other two forms. It is important to realize that IRIS clients will not use all three forms simultaneously. Prior to manipulating network and area data the planner will normally map the area data onto his network and then manipulate all the data in network mode. Alternatively, he might map his network data onto areas and then manipulate the data in parcel mode. A stream model might require point and area data to be mapped into network mode. The stream modeler would map point effluent sources onto network nodes and calculate non-point effluent sources (e.g. from urban or agricultural runoff) from land use data attributes of the parcels adjacent to the links in his network. Once IRIS has mapped both area and point data into network mode, the modeler may proceed with his analysis.

When manipulating area data, IRIS must provide access to a separate data base for each level of geographic resolution for two reasons. First, it is less expensive to maintain aggregated data in a coarse level data base than to continually aggregate fine resolution data to a coarser level. Second, much data is applicable only to a single level of resolution: for example, the number of federal programs in a county is a property of a county and is not applicable to a section, river basin, or other data base.

Figure 2 shows the levels of geographic aggregation which the initial IRIS system will be capable of manipulating. The system will



LEVELS OF GEOGRAPHIC RESOLUTION

FIGURE 2

also have the capability of including new data bases at other levels of geographic resolution as needed.

Time series data is used in air quality analyses at a fine resolution. Census data has a much coarser time resolution. Initially, time can be recorded as an attribute of data stored on a point, area, or network link. This is adequate for most census requirements but marginal for handling fine resolution air data. In the latter case, the modeler will frequently access geographic data as an attribute of a given time rather than time as an attribute of geographic data. Research is required to efficiently integrate time series data into the languages, data bases, and access algorithms of a geographic information system.

Many state files have one or more disclosure constraints. In order to make IRIS data as useful as possible there should be as few security constraints as possible. An averaging and aggregation technique has been developed to present aggregations of such data in a disclosable form [1]. Before it is entered into the system, data is transformed into a form that is suitable for disclosure, and disclosable averages are provided for non-disclosable cells. This technique provides greater privacy than is currently provided for census tapes and it may make more data available for planning purposes. The technical feasibility and desirability of providing greater security, through password protection, is questionable at this time.

The data base should be structured as a multi-file rather than as a single file data base. Operational experience will enable a data base manager to group together those data frequently used concurrently into a single file. For example, one would normally expect to find soils and geological data in the same data base file but might reasonably expect to find census data in a different file. The purpose of the multifile data base is to improve response time and decrease

costs. The multifile distribution of data should be invisible to the user.

The IRIS data base will be very large. The State roadway file alone occupies several hundred million bytes in compressed form. Since the data bases are so large and since a large amount of data will normally be searched in a single request by a single user, both data updating and entry should be considered secondary to retrieval and analysis. In general, retrieval will be on a minute-to-minute basis in an interactive fashion, but storage and updating of data will occur infrequently and will not normally be done by the user.

The requirement for rapid access to the IRIS data base has important hardware and software ramifications. Consider, for example, a NARIS II request to form a study region by examining quarter-quarter sections in a six-county region. If a multifile data base were used to reduce record size, and if all data for this request were in one file and maintained on a disk pack system with IBM 2314 characteristics, it would take at least 30 seconds just to read the data and possibly 5 minutes if the records were poorly allocated on the disk pack. The load on the central processor becomes greater and response time is degraded when multiple file requests are made or when several users are simultaneously exercising the system.

Most of the data management load on the system is caused by generating new keys to the data base (context searches) and by forming the unions, intersections, and exclusions of keys. None of these operations requires a powerful central processor. A properly constructed programmable data channel, similar to a Control Data peripheral processor, could create and manipulate these information keys with minimal intervention by the central processor. Since these processes are inherently parallel, one can achieve an arbitrary reduction in response time by adding the appropriate number of intelligent channels.

In the geographic data analysis problem, the majority of computer time is spent performing a very few operations on information keys. Moving these operations to an intelligent channel drastically reduces main processor overhead and permits many simultaneous context searches witout requiring additional main processors. It also improves the user's ability to tune his hardware configuration to a specific mix of complex analysis and key manipulation activities.

Parker [4] detailed the hardware and algorithms required to perform parallel context searches in a large data base. The Data Computer [5,6] being built by the Computer Corporation of America takes a different approach. The Data Computer, a single large processor (PDP-10) attached to a trillion bit store (Unicon 690), performs all of the high and low level data management activities normally executed in a central processor. The intelligent channel is less powerful than the Data Computer yet capable of more global control than the logic-per-track system proposed by Parker.

Research funds have been requested from the National Science Foundation to purchase suitable hardware to test the intelligent channel concept.

Interface Packages

It is not technically feasible to design a single problem solving language suitable for the range of activities and users encountered in the State of Illinois. The IRIS system must reflect this by providing separate languages and interfaces appropriate to the unique jargon and needs of these problem solving activities.

Costs and problem constraints require that IRIS be a conversational system. Many of the questions posed to IRIS cannot be answered by the provision of a single batch program. Rather, they require exploratory access to one or more data bases. The investigator

will interrogate the data base, use the information obtained from the first request, immediately formulate a follow-on request, etc., until the information required to solve the problem has been obtained. For example, the Northeast Illinois Natural Resource Service Center estimates that 200 people each day in the State of Illinois ask their Soil and Water Conservation District, their Cooperative Extension Agent, or some other county agency why their well has run dry. It is not technically feasible to write a single computer program that will interrogate the data base and answer this question without human assistance. However, it is feasible to have the problem solver conversationally interrogate the geological and land use data bases to answer the question and identify corrective actions.

The information contained in the IRIS data base must be made available to the problem solver for his immediate use, in much the same manner he currently uses maps and reference works kept on file in his office. The cost of providing only one remote batch terminal for each county is in excess of two million dollars ($2,000,000) annually. Better access to the data base could be obtained from a conversational terminal at less than one tenth the cost of batch terminals. The cost of batch terminals would discourage the use of the system and prevent the needed exploratory access to the data base.

The interface packages to the IRIS data base must serve both the day to day problem solver and the developer of new analytic techniques. Some interface package design constraints are indicated in order of importance:

1. Any interface package should be capable of using any other. For example, a NARIS II user might need statistical capabilities within the NARIS II language. NARIS II could log into the Statistical package and prepare a statistical request without the NARIS II user knowing it had been done. The Statistical system would

view the NARIS II formulated request as if it had come from any other user who might be logged into the Statistical system from a remote terminal.

2. While the greatest emphasis in interface language design is to make common analyses as easy as possible for the unsophisticated, single package users, the language should still be as similar to other interface languages as possible. Parallelism in the languages will aid the sophisticated multipackage user.

3. Common interface packages are normally used by problem specific interface packages, not directly by the user. Therefore, an interface language for common packages can be oriented towards the computer scientist who writes the problem specific packages.

<u>Program Certification</u>

Software modules within the system should whenever possible, be implemented using system correctness techniques. The current state-of-the-art in large software package design allows a programmer to prove that his program is free of all software errors. The basic correctness proof is performed by exhaustively driving the software system into all relevant states and testing for correct operation. A hierarchical program structuring technique is used to reduce the number of relevant system states which must be tested even though the number of possible states into which a system can be driven still remains very large. This makes it logistically possible to test the system by exhaustively examining a few hundred states rather than billions or trillions. Several large codes have been successively built in this way. Dijkstra pioneered the techniques with the "THE" multi-programming system [7] in Holland. In this country some language compilers [8] and some of the

initial portions of the ILLIAC IV operating system [9, 10] were developed using these techniques.

Concern with program correctness is more than academic. Planning projects are inherently controversial and the analyses performed are frequently challenged in court. While the discussion of analysis techniques is deferred to expert witness, it is possible to rule the results of a computer generated analysis inadmissible if there is strong probability of an error in calculation.

A code certification group must be created which is responsible for testing and certifying IPIS computer codes. Six levels of certification are defined in order of increasing probability of program error:

<u>Correct code certified on a correct operating system</u>: This is the highest level of certification. In order to be certified "correct on correct" a code must be rigorously (in the mathematical sense) proven correct when operating on an operating system which has itself been proved correct.

<u>Correct code certification on a reliable operating system</u>: In order to be certified "correct on reliable" a code must be rigorously proved correct when running on a "reliable" operating system. The definition of "reliable" is critical. While such a code is not rigorously correct, because operating system malfunctions are still possible, the code should still be extraordinarily free of errors in calculation.

<u>Dual code certified</u>: A particular analysis performed by a computer program is dual code certified if two programs, each written to the same specification by different groups and with no known errors in either code, produce identical results.

Expert certified: A code is expert certified if it has no known errors and "experts" who were not involved with the generation of the program examine the logic of the program in detail and offer an expert opinion that the logic is correct.

No known error certified: A code is certified to contain no known errors if it has been exercised with sufficiently diverse input over a significant period of time under controlled test conditions without a failure being discovered.

Uncertified code: Code is uncertified if it cannot be placed in any of the above levels of certification.

Due to the level of skill required to design correctness proofs and the scarcity of software engineers with those skills, most of the system will have to be certified using dual code, expert, and no known error techniques.

IRIS code will be developed in four phases. Phase 1 is the detailed design phase. In phase 2, that design is documented with flow charts and a preliminary user manual is prepared by a technical writer. Phase 3 is the actual coding and debugging. Phase 4 is dedicated to the preparation of a finished user manual and program logic manuals to be used by software engineers for program maintenance and modification.

The IRIS code certification group will become active at the end of phase 1. They must examine the design and flow charts to determine at what level the code can be certified. If not acceptable, design alterations will be indicated to permit easier or higher certification. At the end of phase 3, the code certification group is responsible for checking correctness proofs and testing codes to determine the level of certification to be awarded to the code.

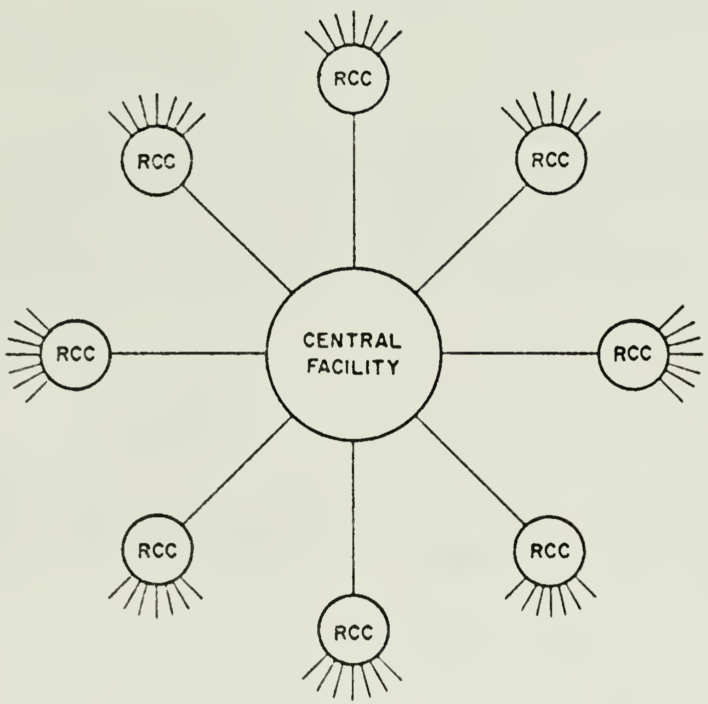
Hardware

Figure 3 illustrates a suitable network configuration for IRIS in the State of Illinois. IRIS will run on a central facility in the State connected by telephone lines to each county. Eight strategically placed remote concentrator computers (RCC) are included in the configuration.

A local planner will dial a nearby RCC rather than the central facility. The RCC's will reduce telephone line costs and provide a small nearby computer, tied into the IRIS central facility, which is capable of operating a local map generator.

The central facility is initially a one million instruction per second (1 MIP) processor with 800 million bytes of direct access, disk pack storage. This must be expandable to a 2 MIP processor with 1,600 million bytes of direct access, disk pack storage. Architectures compatible with the intelligent channel concept are preferred. For reasons of reliability, emphasis should be placed on twin processor configurations and other computer architecture features which allow continued operation at a degraded level in the event of a component failure. Preliminary analyses of Burroughs, CDC, DEC, Honeywell, IBM, NCR, Univac, and XDS hardware and software indicate that no vendor can supply all of the capabilities which would be ideal for IRIS implementation. Furthermore, each of the eight vendors has unique software and hardware facilities which would be valuable for IRIS development and operation. There is no obvious sole source candidate. A bid procedure with a carefully designed analysis of bids will be required to choose an optimal vendor.

Hardware related operation costs are tabulated in tables 1 and 2 for the central facility, remote concentrator computers, telephone lines, and personnel. Maximum and minimum ranges on central facility costs are provided.



(RCC = REMOTE CONCENTRATOR COMPUTER)

IRIS NETWORK
FIGURE 3

Central facility costs are computed using the list prices of IBM and DEC equipment. These two vendors were chosen for these estimates because they tend to represent the upper and lower bounds of expense found in the open market. Costs are estimated using list prices for approximately equal configurations. Table 1 summarizes the configurations and costs.

		Processor speed in millions of instr. per sec.	Number of processors and type	Core size in millions of bytes	Disk size in millions of bytes	Estimated purchase in 1000's	Estimated yearly lease in 1000's
initial	IBM	1/2	1-370/155	1.5	800	$2909	$ 775
	DEC	1	1-KI10	1.2	800	$2417	$ 610
expanded	IBM	2	1-370/165	2.0	1600	$5048	$1308
	DEC	2	2-KI10	2.1	1600	$3002	$ 983

Initial and Expanded Central Facility Configurations
and Cost Estimates

Table 1

Remote concentrator computer costs are estimated for a minimum and maximum RCC. A minimum RCC is a processor such as a PDP-11/20 with 10 dial up lines and an interface to a leased line to the central facility. A maximum RCC is a processor such as a PDP-11/45 augmented with an electrostatic plotter and a 45 million byte disk pack (required to do IRIS maps over an inexpensive 2400 baud leased phone line). A minimum RCC costs approximately $25,000 and a maximum RCC costs approximately $107,000. In the initial configuration 6 minimum RCC's are included. In addition, two maximum RCC's, with graphics capabilities, are provided - one for production use at the central

facility and one for use at the user training facility. The expanded configuration provides 8 maximum RCC's.

Eight leased lines at 2400 baud with an average length of 125 miles is assumed for telephone line costs (approximately $11,000/yr.) In addition ten dial up lines for each of eight RCC's are included approximately $11,000/yr.).

Table 2 summarizes the yearly costs incurred in maintaining the software and renting the hardware. Personnel and indirect costs are calculated by doubling all hardware related costs. This will include operations, management, and software maintainance staff. This does not include software development staff for major enhancement efforts.

	Central facility DEC-IBM	RCC's and Graphics	Phone lines	Personnel and indirect costs DEC-IBM	Total DEC-IBM
initial	610-775	92	22	724-889	1448-1778
expanded	983-1308	216	22	1221-1546	2442-3092

Yearly Hardware Related Costs in Thousands

Table 2

Common and Problem Specific Interface Packages

The following discussion provides additional information about the seven interface packages proposed for the initial IRIS system. Five of these are common and two are problem specific. All interface packages share the same files. Thus interface packages can and will prepare and process files for other interface packages. Problem specific interface packages address specific problem areas and have user languages suitable for use by problem solvers in those areas. Common interface packages provide support and data manipulation facilities that are common to two or more IRIS interface packages. They also provide the common facilities, such as data base description, that are needed by all users. The common interface packages will be accessed through problem specific interface packages rather than directly by the problem solver. The two problem specific and five common interface packages are described below.

NARIS II

The NARIS II interface packages to IRIS should provide those facilities currently available to the planner through NARIS [2,3] NARIS is described in chapter 4.

NARIS will be extended to handle data bases at multiple levels of geographic resolution. Furthermore, an ability to manipulate network data and describe networks or groups of points in much the same way that NARIS can now describe regions of quarter-quarter sections for study will be added. The current NARIS system allows the users to create only one level of region, a study region which includes only tracts. It is desirable to create sub-regions, sub-regions of sub-regions, etc. Then a study region can consist of tracts or sub-regions. This facility will enable the user to prepare partial analyses on special sub-regions and readily include these partial analyses in the analysis of a larger region analysis. It will also significantly enhance the report

generating capabilities of NARIS by providing natural sub-units for data summaries and distributions within regions.

Statistical System

Social and regional planning agencies need to extract data from the IRIS data base and then manipulate it through standard statistical tests: correlation, regression, multi-variate analysis, etc. An interface package to the IRIS data base with its own statistics oriented interface language will be provided. It will allow the user to manipulate the basic IRIS data base, or a modified data base prepared by, for example, the NARIS II package, in a statistical environment.

File Exporter

As indicated earlier, an interface is required that would generate files compatible with other software packages. Those software packages could be on the IRIS machine or on a different machine. The file exporter will initially consist of many single purpose programs that will read an IRIS file and write it into a specified format. As experience is gained with writing many files, a detailed analysis of the file exportation experience will determine the canonical operations. It is expected that this will produce a very small group of programs with bread and powerful file reformatting capabilities.

Graphics

A graphics package is required to perform the mapping functions of NARIS II and the graph generation functions of the Statistical system. Electrostatic and line printer graphics will be supported.

General Purpose Report Generator

The general purpose report generator is required as support for all other interface packages which generate tabular summaries. A cost saving can be achieved if one general purpose report generator is provided. The problem specific interface packages will receive report generation requests from users in their own particular user specific languages, map these into requests for the general purpose report generator, and then transmit a data file to the general purpose report generator. The report generator will then produce the tabular summary.

Data Dictionary

A Data Dictionary with a powerful cross-indexing capability must be provided. This interface package will allow the user to determine what data is available in IRIS. Documentation will be stored on each type of data describing its date of collection, accuracy, generating agency and other pertinent information.

Data Entry

Data entry will require the greatest amount of programming effort. Experience with other systems has shown that the data prepared for these systems frequently comes to the computer center in a wide variety of formats and ready for input containing many obvious and not so obvious errors. Special purpose data cleaning programs must be designed for each specific data type which will examine the data fields of completed recording forms for legitimacy.

The data entry group will be responsible for preparing the digitizer software required to process mapped data.

References

1. "Disclosure and Data Base Security", IRIS design memorandum #1, Center for Advanced Computation, University of Illinois, Urbana, November 12, 1971.

2. "NARIS: A Natural Resource Information System", Center for Advanced Computation, University of Illinois, Urbana, 1971.

3. McTeer, W., et. al., "NARIS User's Manual, Center for Advanced Computation, University of Illinois, Urbana, February 1971.

4. Parker, J. L., A Logic Per Track Information Retrieval System, Ph.D. Thesis, ILLIAC IV Doc. #242, University of Illinois, Urbana, February 1971.

5. "Datalanguage", Data Computer Project Working Paper No. 3, Computer Corporation of America, Cambridge, October 1971.

6. "Timings and Storage Estimates", Data Computer Project Working Paper No. 4, Computer Corporation of America, Cambridge, November 1971.

7. Dijkstra, E.W., "The Structure of the 'THE' - Multiprogramming System", CACM, Vol. 11, No. 5, May 1968, pp. 341-346.

8. London, R. L., "Correctness of a LISP Subset", SIGPLAN Noti________, Vol. 7, No. 9, Jan. 1972, pp. 121-127.

9. Alsberg, P.A., and Mills, C.R., "The Structure of the ILLIAC IV Operating System", Second Symposium on Operating Systems Principles, Princeton University, October 20-22, 1969.

10. Thomas, J., A B6500 ALGOL Program, Vols. 1 and 2, ILLIAC IV Project, University of Illinois, Urbana, August, 1971. p85

CHAPTER 6
COSTS AND SCHEDULES

The problems and circumstances which precipitated the need for IRIS are not unique to Illinois. Other states are subject to the same federal guidelines and are experiencing planning information problems similar to our own. Some of the states have attempted, independently, to develop geographic information systems to meet their own needs. Their experience has shown that small, special purpose information systems soon become obsolete and cannot be readily adapted to meet changing societal needs. On the other hand, a large versatile system such as IRIS can only be developed through an intensive three year program involving several major advances in the state of the art in geographic information systems.

Due to the innovative nature of the program and the nationwide need for its products, it is recommended that the State and the University jointly seek funding from an appropriate federal agency. The Research Applied to National Needs (RANN) program at the National Science Foundation has already been contacted. The Ford Foundation, which sponsored the development of the Natural Resource Information System (NARIS) at the Center for Advanced Computation is interested in the IRIS program and will discuss its funding interest with the National Science Foundation.

A three year research and development budget is presented on the following pages. The costs are separated into federal and state shares. It is recommended that federal funding be sought to cover the cost of all research and the costs of interpreting, assembling, and encoding existing data for the statewide data base. Thus, federal funds should pay for the development of the computer system and inputting an initial data base to prove the concept. In addition, federal funds are requested for the collection and encoding of fine resolution data for

the Fox river basin. This data base is experimental in nature and can be used by researchers to determine the cost-effectiveness of collecting data at fine vs. coarse geographic resolutions.

It is recommended that the State provide for the administration of the program within the government. It is also recommended that the State collect any new data required for the statewide data base. The need for the following new data was frequently expressed to IRIS project personnel during the survey of potential users. First, information on current land use should be obtained from aerial photographs at quarter-quarter section resolution over the entire state. Approximately thirty land use classifications can be readily detected from 1:24,000 aerial photographs. This will produce an excellent initial land use data bank. Second, the location of natural and ecologically unique areas should be determined by the State and input into the initial statewide data base. Much of this information already exists in universities, the Natural History Survey, and the State Museum. It should be assembled and made available to State agency personnel for environmental impact studies. The need for this information does not depend on IRIS. The State should acquire this information as soon as possible. Concurrent development of IRIS only adds to this urgency.

SUMMARY BUDGET FOR RECOMMENDED THREE YEAR IRIS PROGRAM
(includes computer system development and initial data base acquisition)

Federal Funds

	1973	1974	1975	TOTAL
Salaries & Wages	$ 464,255	$ 508,155	$ 417,825	$1,290,245
Fringe Benefits	57,655	62,660	51,560	171,875
Indirect Cost	292,480	320,140	263,240	875,860
Wages and Related Costs	814,390	890,955	732,635	2,437,980
Computer Time	155,520	257,040	158,400	570,960
Travel	21,250	19,500	18,000	58,750
Equipment Leases, Supplies and Maintenance	92,240	181,860	240,940	515,040
Equipment Purchased	68,000	104,000	18,000	190,000
Publications	2,400	8,000	8,000	18,400
Data Preparation and Encoding	230,325	414,585	276,390	921,300
Consulting Services for Clients	55,000	57,000	58,000	170,000
General Expense (incl. Xerox, telephone, books, etc.)	103,600	113,100	99,675	316,375
TOTAL	$1,542,725	$2,046,040	$1,610,040	$5,198,805

State Funds

	1973	1974	1975	TOTAL
New Data Acquisition	$ 265,000	$ 260,000	$ 17,000	$ 542,000

DATA COSTS

Federal Funds

Statewide Data Base

DATA TYPE	COST
Population and Housing	$ 19,700
Health, Education, Welfare, Public Safety, and Recreation	11,000
Economics	23,000
Geology	5,500
Soils	245,000
Agriculture	6,500
Transportation	140,000
Historic, Architectural & Archaeological Sites	5,000
Air	1,600
Water	34,000
Total	$ 491,300

Fine Resolution Data Base

DATA TYPE	COST
Geology for Planning	$ 122,000
Soil Types	16,400
Native and Planted Vegetation	11,000
Impoundment Sites and Small Watersheds	19,600
Detailed Land Use Survey	240,000
Assessment Data	5,500
Extension of NIPC Socio-Economic Data Coverage	5,500
Refinement of NIPC Data to 1/4 1/4 Sections	10,000
Total	$ 430,000

State Funds

Statewide Data Base

DATA TYPE	COST
Land Use	$ 515,000
Natural and Ecologically Unique Areas	27,000
Total	$ 542,000

SCHEDULES

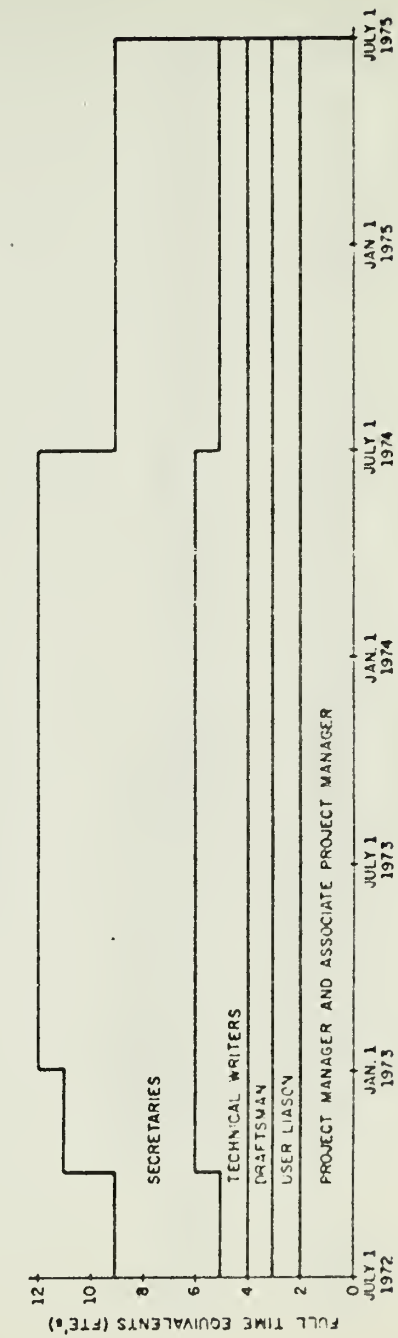
A three-year computer system program is recommended. It consists of six software tasks:

1. Software Certification
2. Nucleus and Remote Concentrator
3. NARIS II Interface Package
4. Statistical Interface Package
5. Data Entry
6. Common Interface Packages

Each of the six software groups should be staffed by professional software engineers, graduate research assistants, and one senior software engineer who will be responsible for task completion.

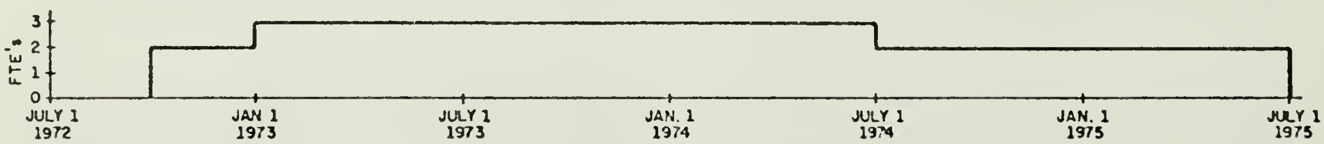
A full time User Liaison should be employed to facilitate data flow to the University in the first year and prepare and present training lectures to State and other interested users in years two and three.

After 18 months a working minimal system will be available for an initial set of clients. Subsequent refinements and enhancements will produce a generally useful system, loaded with a working data base ready to be placed in a production environment at the end of the three year program.



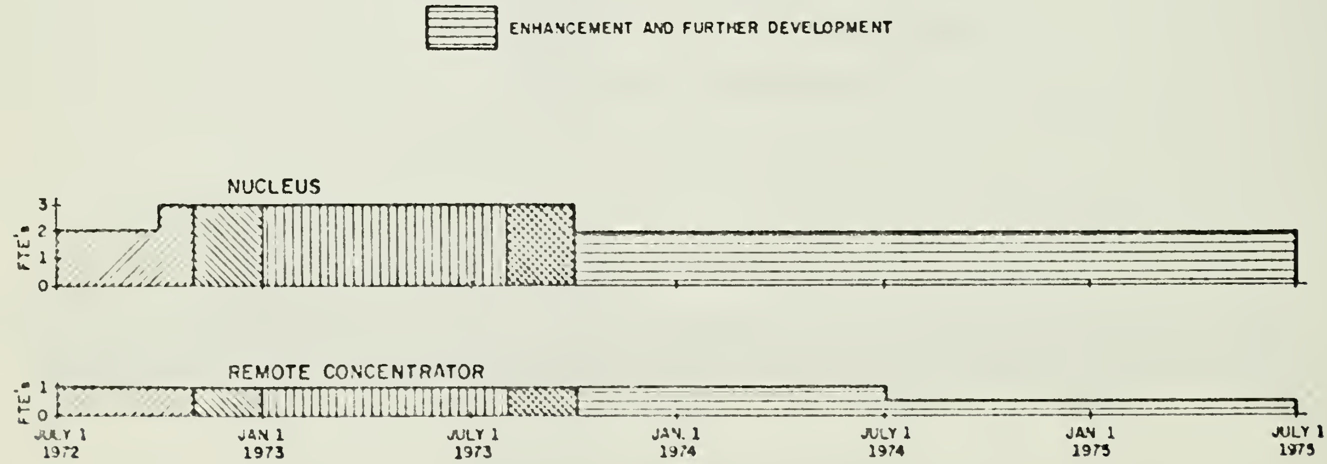
IRIS STAFF SUPPORT PERSONNEL

FIGURE 1



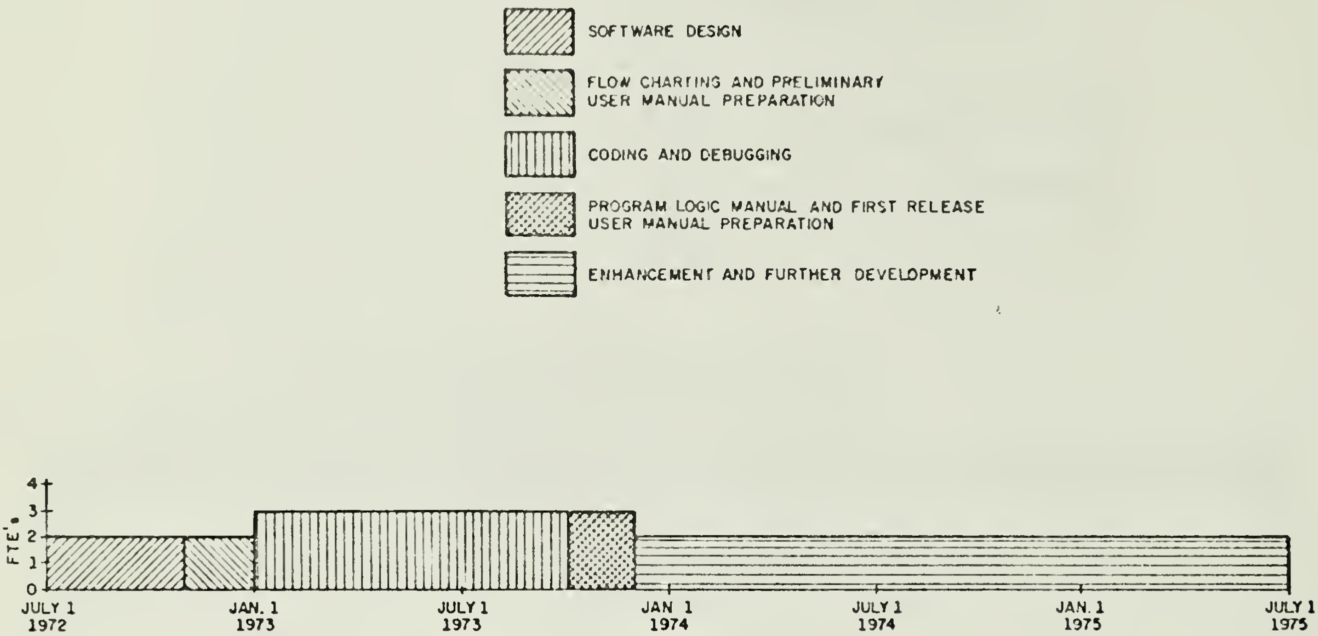
SOFTWARE CERTIFICATION

FIGURE 2



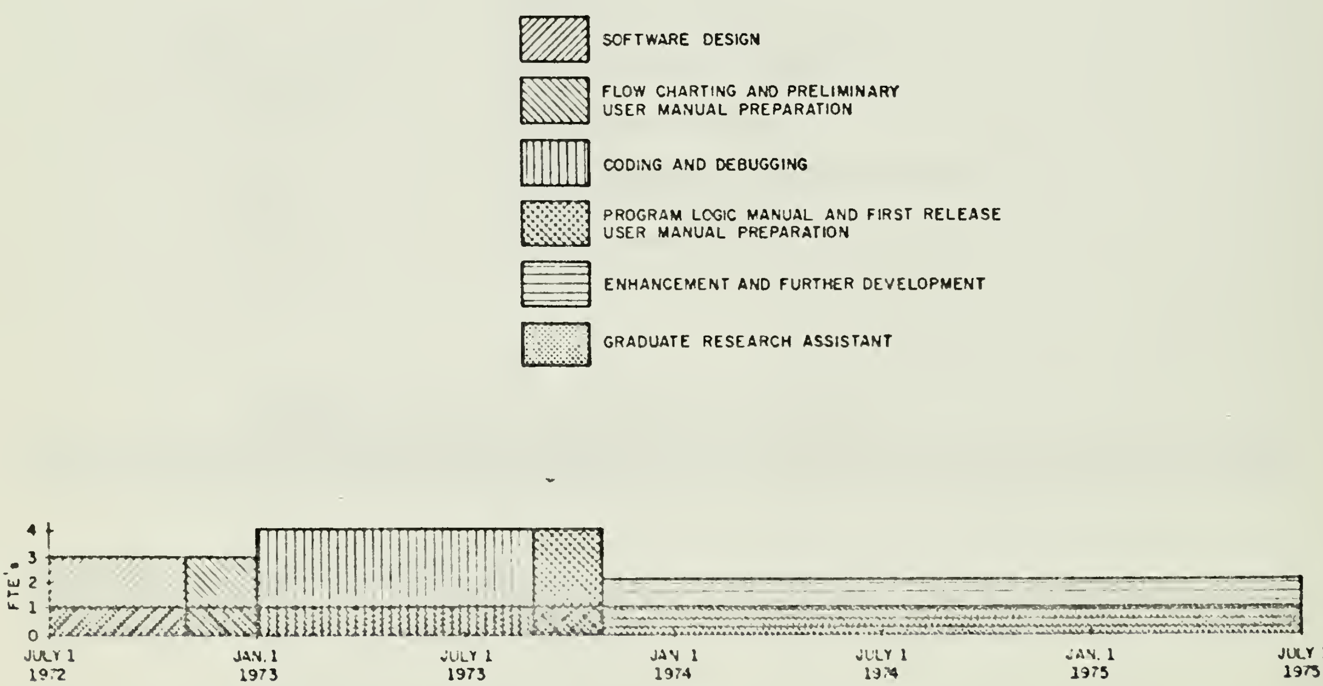
NUCLEUS AND REMOTE CONCENTRATOR

FIGURE 3



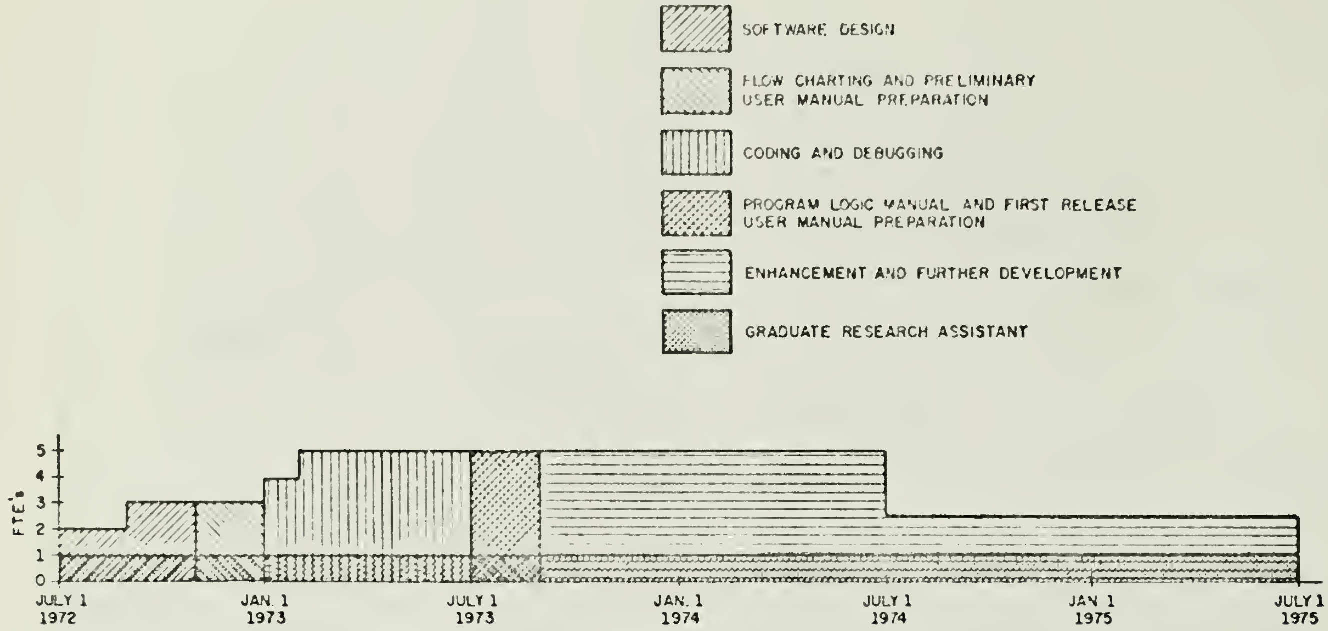
NARIS II INTERFACE PACKAGE

FIGURE 4



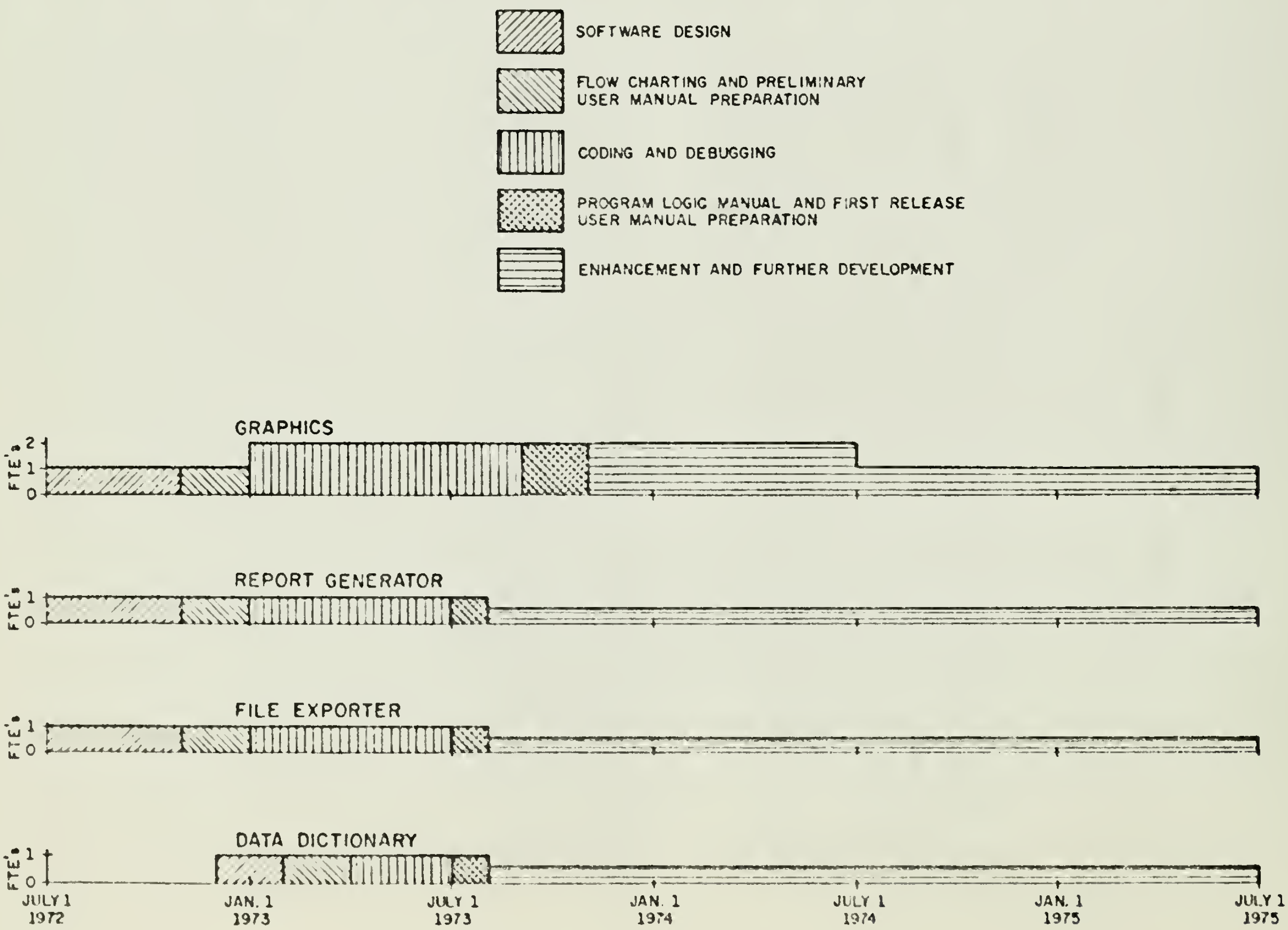
STATISTICAL INTERFACE PACKAGE

FIGURE 5



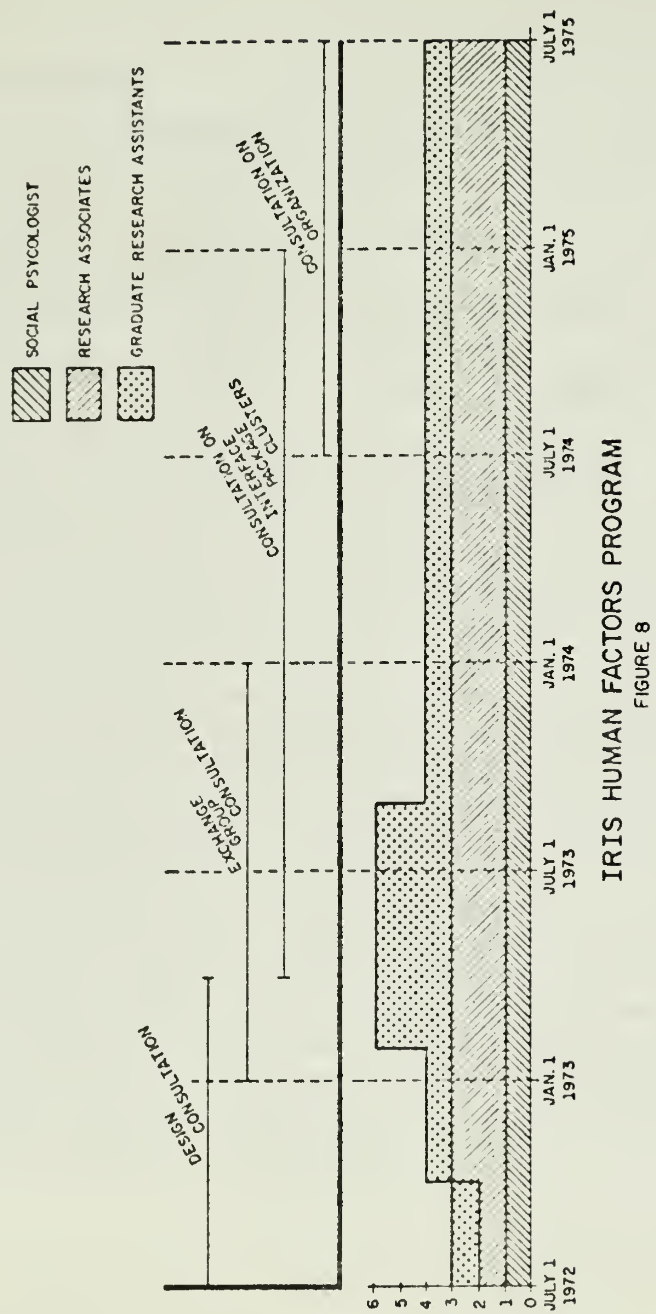
DATA ENTRY

FIGURE 6



COMMON INTERFACE PACKAGES

FIGURE 7



IRIS HUMAN FACTORS PROGRAM

FIGURE 8

APPENDIX A

INITIAL DATA BASE RECOMMENDATIONS

Statewide Data Base

The determination of statewide priorities for land use, transportation, recreation, and other activities requires access to a data base covering the entire state. The initial data base should be assembled largely from existing data. The system should be designed to accept existing geographically referenced data at resolutions ranging from precisely located points to county level aggregations. Indices should be constructed to allow users to select a geographic resolution at which to work and to give them access to all data input at finer geographic resolutions.

In the following tables, the names and source agencies are listed for all files which should be included in the initial statewide data base. The geographic resolution at which data are currently available is specified for area-related data. Most point and network data have geographic resolutions of approximately 100 feet, well within the smallest parcel size, the quarter-quarter section.

Population and Housing

Census of population and housing	County, Municipality, Tract	U.S. Bureau of the Census
Vital statistics	Township, Municipality	Department of Public Health
Population estimates	Township, Municipality	Department of Public Health
Population forecasts	County, Municipality	Department of Business and Economic Development
Construction Reports	County, Municipality	U.S. Census Bureau, Illinois Office of Planning and Analysis

The Survey Research Laboratory at the University of Illinois will receive the data from the 1970 Census of Population and Housing on magnetic tapes from the Census Bureau. The University of Illinois is fortunate in that its Survey Research Laboratory is developing the capability of servicing requests for processed census data. The existence of this facility will minimize the cost of reading these data into the IRIS data base. The initial statewide data base should include census data at the county, municipality, county subdivision, and tract levels of geographic resolution. Additional population and housing data will be obtained from several State agencies. The Illinois Department of Public Health calculates monthly and annual estimates of current population for each incorporated township from vital statistics and school enrollment data. Population forecasts for counties and municipalities are assembled by the Illinois Department of Business and Economic Development from a variety of primary sources. Finally, construction reports from the U.S. Census Bureau are received by the State Office of Planning and Analysis; information from these reports could be used to update land use and housing information.

Health, Education, Welfare, Public Safety, and Recreation

Health and safety facilities	County, Municipality	Department of Business and Economic Development
Day care and health facilities	County, Municipality	Office of Human Resources
Public health statistics	County, Municipality	Department of Public Health
Mental health statistics	County, Municipality	Department of Mental Health
Educational enrollment	County	Superintendent of Public Instruction
Public aid statistics	County, Municipality	Department of Public Aid
Crime statistics	County	Department of Law Enforcement
Inmates, discharge, and parole	County	Department of Corrections
Firearm owners	County, Municipality	Department of Law Enforcement
Hunters and fishermen	County	Department of Conservation
Motorboat owners	County	Department of Conservation

A considerable amount of data already exists in machine readable form. In fact, the Illinois Department of Public Health is constructing a Total Health Information System (THIS) for their own internal administrative data processing. With the cooperation of the Department, data on morbidity, mortality, and licensed health facilities could be read into the IRIS data base at minimal cost. In the Departments of Mental Health, Corrections, and Public Aid, most data of fine geographic resolution are maintained in confidential files and only aggregated data of coarse geographic resolution is available. With the cooperation of these Departments it would be possible to use such files to prepare disclosable data of a finer geographic resolution for the

IRIS data base. In addition to the county and municipality geolocators, ZIP codes are available for most data items.

Recreation data of a geographic quality compatible with IRIS is relatively unavailable at this time. As Illinois' first comprehensive recreation plan is developed, the Department of Conservation may obtain more data and make it available to the IRIS project.

Economics

Censuses of business, manufacturers, and governments	County, Municipality	U.S. Bureau of the Census
Property taxes	County, Township, Municipality	Department of Local Government Affairs
Public Finance	County, Township, Municipality	Department of Business and Economic Development
Business taxes	County, Municipality	Department of Revenue
Elements of the economic base	County, Municipality	Department of Business and Economic Development
Public utilities	Digitized networks and areas	Illinois Commerce Commission
A-95 data	County, Municipality	Office of Planning and Analysis

The U.S. Departments of Commerce and Labor are primary sources of economic data for Illinois, most of which are available only at the county level. Other information for smaller geographic areas within the State is available from the Illinois Departments of Local Government Affairs, Revenue, Labor, and Business and Economic Development. With the cooperation of personnel from these Departments, the primary data could be assembled and read into the IRIS data base. Most of the information is currently available in machine readable form, but some of the annually updated fine resolution data will have to be encoded. The Illinois Commerce Commission is charged with regulating all public utilities in the State, and has maps and other data on railroads, pipelines, electric generation facilities and transmission networks, etc. Finally, the Office of Planning and Analysis is the Illinois clearinghouse for the federal A-95 program. Under this program, federal

projects are evaluated for their compatibility with regional and State plans. In the course of these program reviews, a considerable amount of physical, social, and economic data relating to the proposed programs should be channeled through this central clearinghouse and converted to a form compatible with IRIS.

Geology

Surface deposits	Digitized area	State Geological Survey
Oil, gas, and water wells	Point	State Geological Survey
Oil and gas fields and storage	Section	State Geological Survey
Oil and gas statistics	Township	Department of Mines and Minerals
Coal statistics	Point	Department of Mines and Minerals

Two state agencies maintain geological data. The Illinois State Geological Survey has produced maps of the surface deposits for the entire state and maintains locational information on oil and gas fields and storage areas. The Department of Mines and Minerals produces annual statistics on producing oil and gas wells and coal mines. The State Geological Survey is currently engaged in a $70,000 program to encode drilling data for over 200,000 water, oil and gas wells in the State. This geological information will be compatible with IRIS and could be included in the data base as soon as it is available.

Soils

Soil associations/ interpretations	Section	University of Illinois
Soil types	County	USDA Soil Conservation Service
Soil productivity	County	University of Illinois
Soil and water conservation needs	County, Watershed	USDA Soil Conservation Service

Sufficient soil survey information is available to determine and encode the soil associations present in each survey section throughout the State. It will be necessary for soil scientists to interpret existing soil survey data to provide information with consistent nomenclature for the entire state. A description of limitations and suitabilities for selected uses would be included for each soil association in the data base. As a result of a two percent sampling program undertaken recently by the Soil Conservation Service, information on the occurrence and productivity of each soil type is available for every county. These data, together with county information on soil and water conservation needs, should also be included in the IRIS data base.

Agriculture

Census of Agriculture	County	U.S. Bureau of the Census
Annual farm census	Township	Illinois Department of Agriculture
Grain production and storage	County	University of Illinois
Fertilizer sales and use	County	Illinois Department of Agriculture/USDA
Agriculture product prices	Crop Reporting District	Illinois Department of Agriculture/USDA
Pesticide/herbicide use	Crop Reporting District	Illinois Department of Agriculture/USDA

Agriculture is the largest single industry in Illinois. In addition to the U.S. Census of Agriculture (conducted every five years) the State of Illinois conducts an annual farm census. Data gathered in this annual census are aggregated and published for each incorporated township. The U.S. census provides only county data. Other county level data collected by the Illinois Department of Agriculture and the U.S.D.A. Cooperative Crop Reporting Service includes information on grain production and storage, agricultural product prices, and sale and use of fertilizers, pesticides and herbicides. The availability of such data in a geographic information system would be useful not only to agricultural planners, but also to planners in the fields of economics and water quality management and researchers investigating pesticide and fertilizer pollution.

Land Use

Aerial photography	1:24,000	Private contractor
Land use interpretation	1/4 1/4 section	Illinois Institute for Environmental Quality and the Office of Planning and Analysis

The statewide IRIS data base must include a consistent set of land use information. Since there has been no need for statewide uniformity in land use codes in the past, existing data is only of marginal value. It is therefore necessary to obtain a complete set of aerial photographs for the entire state. The photographs should be interpreted for only about 30 types of land uses in order to minimize or eliminate the need for expensive ground truth checks. Then, using existing data files from the Bureau of the Census, the Illinois Department of Labor, and the Environmental Protection Agency, lists of manufacturers and industries should be assembled for each county and SIC classification. The locations of these facilities could then be determined within each county by the Illinois Office of Planning and Analysis, resulting in a detailed land use classification system. Procedures for updating this data using IRIS should also be developed.

Transportation

Road file	Network	Department of Transportation
Community transportation facilities	Municipality	Department of Business and Economic Development
Census of transportation	State	U.S. Bureau of the Census
Airport information	Point	Department of Aeronautics
Motor vehicles	Township, Municipality	Secretary of State
Aircraft	County	Department of Aeronautics
Pilots	County, Municipality	Department of Aeronautics

Illinois is fortunate in having most of its data pertaining to highway transportation in a single computer file at the Division of Highways. The geographic locator for the file, however, is the road mile index, and the tape file is so large and difficult to handle that it is read completely only once per year to allocate tax revenues. A major reason for not using this file more frequently is the difficulty of correlating its network data with area-related information. It is therefore proposed that the highway network be digitized using existing county highway maps which show all federal, state and county roads. Virtually all the network data from the road file, including traffic counts, accidents, and physical data such as highway and right-of-way width, intersections, and structures, could be made available through IRIS in such a form that it may be manipulated with other socio-economic and natural resource data in the system. Other information for transportation planning such as data on community transportation facilities, airports, aircraft, pilots, motor vehicles, and data from the U.S. Census of Transportation should also be included in the data base.

Wildlife Resources, Natural Areas, and Historic Sites

Natural areas	1/4 1/4 section	Department of Conservation, State Natural History Survey, State Museum, Universities, etc.
Wildlife occurrence	Point, area	as above
Historic sites	Point	State Historical Survey
Archaeological sites	Point	State Archaeological Survey
Architectural sites	Point	State Architectural Survey

Most existing data on natural areas and wildlife resources in the State are of a geographic quality that is at best inconsistent. For the first time, the Department of Conservation is compiling a statewide inventory of large natural areas which have potential as future State parks or conservation areas. Unfortunately, most information on smaller natural areas and areas of unique ecological significance exists only in the technical reports and files of personnel at the Illinois Natural History Survey and professors at universities throughout the State. In order for IRIS to be useful to planners writing environmental impact statements, information on such areas must be included in the data base. It is strongly recommended that the State assemble an inventory of such sites through discussion with experts throughout the State and prepare this information for the IRIS data base.

In cooperation with the National Park Service and through the generosity of private individuals and foundations, statewide inventories of historic, archaeological and architectural sites are being compiled. Illinois has approximately 10,000 such sites, locations of which are being recorded in such a way as to be compatible with IRIS. This information should be included so that planners using IRIS may be made aware of the locations of such sites.

Air

Climatological summaries	Point	State Climatologist
Wind, humidity, cloud cover	Point	State Climatologist
Freeze date probabilities	Digitized areas	State Climatologist
Ambient air quality and meterological conditions	Point	Illinois Environmental Protection Agency
Air pollutant emissions	Point	Illinois Environmental Protection Agency

Most of the information on air quality, meterology, and climatology is point data. The Illinois Environmental Protection Agency has extensive data on air quality, pollutant emissions, and meterology. The remaining air data could be obtained from the National Oceanic and Atmospheric Administration through the Illinois State Climatologist.

Water

Stream network digitization	Network	U.S. Environmental Protection Administration
Water quality standards	Network	Illinois Pollution Control Board
Pollutant emissions	Point	Illinois Environmental Protection Agency
Surface water quality	Point	Illinois Environmental Protection Agency, Illinois State Water Survey
Groundwater quality	Point	Illinois State Water Survey
Public water supply quality	Point	Illinois Environmental Protection Agency
Floodplains	Digitized area	U.S. Geological Survey U.S. Army Corps of Engineers
Flow duration and low flow frequencies	Point	U.S. Geological Survey
Time of travel	Network	Illinois State Water Survey
Public water supply adequacy	County, Municipality	Illinois State Water Survey

Water data is primarily point and network information. The costly background work for placing this data into IRIS will have been accomplished through the Federal Environmental Protection Administration STORET digitizing program. The State Environmental Protection Agency and Pollution Control Board could supply all of their water quality informatio to IRIS. Other existing data to be included in the IRIS data base could be obtained from the State Water Survey, the U.S. Geological Survey, and the U.S. Army Corps of Engineers. The most costly, but perhaps the most useful of these data, are floodplain maps which should be digitized and made available to planners through IRIS.

Fine Resolution Data Base

The purpose of the fine resolution data base is twofold. First, it will provide planners and researchers with a combination of high quality natural resource and socio-economic data and the capabilities for manipulating it using IRIS. This will promote research in the use of powerful geographic information systems, research which may result in the development of additional problem specific interface packages. Second, a fine resolution data base, containing data relating to census blocks and quarter-quarter sections, will be used to determine the cost-effectiveness of collecting different types of data at varying geographic resolutions.

It is proposed that the fine resolution geographic data base be assembled for the Fox river basin in Northeastern Illinois. Although it covers less than 5% of the area of the State, the basin is typical enough to serve as a pilot river basin for developing basin planning techniques. The basin contains small watersheds which are completely rural, and others which are entirely urban. The upper portion of the river is dammed, and the lower portion is free flowing. It is located in one of the most rapidly developing areas of the State. Although the basin itself touches nine counties, it lies largely within the eight county area for which NARIS was developed. There is also considerable overlap with the six Illinois counties served by the Northeastern Illinois Planning Commission (NIPC). The overlaps with NARIS and NIPC regions will greatly reduce the scale of the additional data collection effort for the Fox river basin.

Natural Resource Data

Approximately 70% of the Fox river basin is covered by the NARIS region. The NARIS schedule for acquiring and encoding data closely parallels that of the proposed IRIS program. Two of the most expensive items of natural resource data are soils and geological data

and interpretations. Fortunately, the Soil Conservation Service has recently completed soil surveys in the counties covering the entire Fox river basin. Furthermore, the Illinois Geological Survey has produced the first two in a series of county Bulletins entitled "Geology for Planning." To date, only two counties have been completed and current plans call for the completion of the NARIS counties by 1974. To cover the Fox river basin, surveys of two more counties would be required. The Geological Survey has already accumulated almost half of the necessary information on these two counties; it is estimated that $120,000 will be required to complete the surveys and prepare the reports.

Data on native woody vegetation and planted woody vegetation will also be available for 70% of the basin in the NARIS data base and it is recommended that coverage be extended to the rest of the basin. Similarly, data on present and potential impoundment sites should be acquired on the quarter-quarter section grid. Finally, watershed boundaries within the Fox river basin should be digitized in order to permit the aggregation and examination of data within these boundaries to study water quality and water resource management.

Socio-Economic Data

Most counties in Northeastern Illinois have adopted a permanent parcel numbering system for use by county assessors. Its purpose is to permit the posting of assessment information without a legal description of each parcel of land. Under this system, each taxable parcel is assigned a ten digit number indicating, among other things, the quarter-quarter section in which it lies. Over 75% of the ownership parcels in the Fox basin will have such designations in the near future, allowing assessment records to be read into the quarter-quarter section data base and aggregated to that level. Tax maps are available for the rest of the area and could be encoded manually.

The Northeastern Illinois Planning Commission has recently contracted with the Center for Advanced Computation to augment NARIS to serve the needs of the urban planner. Initially, NIPC plans to include population, housing and employment data in the system by quarter sections in its six county planning area. The Commission has proposed to experiment with these and other data and assess the adequacy of NARIS analysis capabilities.

We recommend that the State construct a quarter-quarter section socio-economic data base in the Fox basin. In over half of this area, it would be merely a refinement of the NIPC quarter section data base. In the rest of the basin, information from the statewide data base would be refined to the quarter-quarter section level. Since this area is generally rural, the cost will be comparatively low. Thus, the IRIS program will benefit from the NIPC experiences in determining the types of socio-economic data for use in this fine resolution data base.

Finally, detailed land use data should be collected for the Fox river basin. The classification system should be based largely on the HUD/BPR code. The detailed survey itself would provide a ground truth check on the accuracy of the aerial photo interpretation for the statewide land use data. Combining this detailed land use information with socio-economic and natural resource data at a quarter-quarter section resolution would provide a basis for determining its usefulness to different types of planners. The entire fine resolution data base is in fact an experiment to determine the degree of detail and the geographic resolution required for various types of planning activities. The recommended system will have a built-in monitor to record the number of times certain data were accessed, by which types of users, and for which purposes. The data base and system can also be made available to other researchers, thereby eliminating the need for extensive and costly data acquisition efforts in future studies and planning research programs.

APPENDIX B

TABULAR SUMMARIES OF THE STATE OF THE ART SURVEY

This appendix contains two tables describing the organizations surveyed and summarizations of the capabilities of each system, and system component surveyed. Following the tables are one page summaries of the organizational efforts.

Table 1 describes those relevant organizations included in the IRIS state of the art survey. This table includes approximately 60 organizations who have either developed or plan to develop a geographic information system or a component of a geographic information system, or who have released at least preliminary reports on a study of environmental information system requirements. Other organizations were contacted but are not included in this table or in the summaries. Surveyed organizations were excluded from the tables for one or more of these reasons: 1) they were study groups which had just begun to look at a problem but had not yet produced a body of literature; 2) they were groups who had generated a data base for specific problems but had not built a geographic information system or system component to handle that data base; or 3) they had taken a very specific approach to some small part of the geographic information problem which was not relevant to the State of Illinois.

Table 1 describes the project name, project type, the state or country involved and the survey techniques used for each of the relevant organizations. The project name was typically the acronym used for the system, module, or study group. Three project types are indicated: systems (comprehensive geographic information systems), modules (geographic information system components), and study groups. The state or country involved in the organization study is also tabulated. Projects undertaken at national laboratories are credited to the United

States as a whole and projects undertaken at state universities or funded by specific states are credited to those states. The survey techniques used were reading of the literature, pursuing personal contacts of IRIS project staff with knowledgeable people at the indicated organizations, or site visits to those organizations who had interesting or important systems, modules, or study groups.

Table 2 describes 30 systems and modules investigated in the survey and tabulates their basic attributes and capabilities. Great care must be exercised in interpreting Table 2. In general, geographic information systems fall into very many classes and have overlapping, as well as disjoint capabilities. These capabilities of the system may be in a variety of phases of implementation - from a planning phase to a thoroughly implemented, debugged, and working phase. Rather than try to summarize all the details of such development, a general description of the overall development state of the project has been used. As a result, depending upon the interest of individual readers and their own particular judgments, they may view the system as being in a more advanced or less advanced state for their own purposes than the reviewers did. In order to assure uniformity of review of the systems, two IRIS project members independently read each of the files on each of the 30 systems and modules and rated their capabilities according to a jointly agreed upon, consistent scale. Organizations with systems were then contacted to verify the reviewers' interpretations and provide data on the system current as of December, 1971. A more detailed description of each column in table 2 and the values that may appear in each column of the table follows:

<u>Classification</u>: Systems and modules were classified according to their status and type.

<u>type</u>: Two types are indicated: system and module.

Systems are further classified as uniform grid, nonuniform grid, parcel, area boundary, network, or point systems. General purpose data management systems which were used as geographic information systems are also classified as "systems".

Modules are further classified as cartographic printer mapping programs, a data base file of unique structure, or coordinate conversion programs.

status: Status was chosen from one of five different categories; planned, experimental, feasible, developed, or implemented. A definition of each of these terms agreed upon at the UNESCO/IGU First Symposium on Geographical Information Systems has been used in this report and is reproduced here.

Planned systems have manipulation capabilities that are required and are funded for research and which are logical extensions of current, state of the art manipulation capabilities.

Experimental systems have their conceptual state being worked on and various approaches are being tested but final methods are not determined.

Feasible systems have their conceptual work finished and, where applicable, computer code is being written but has not been checked for errors. Also the cost of use is not determined in detail but can be estimated.

Developed systems have their conceptual work finished where applicable and computer code is being written

and is error free to a minimum workable level. These systems are economically acceptable to users but not yet documented or in routine use.

Implemented systems are being used on a routine basis and have either been documented or can be visited and seen.

Geographic Attributes: The addressing, resolution, and coverage of the system are described.

addressing: The addressing system may be UTM coordinates, x,y coordinates, public rectangular land survey, census blocks and tracts, city blocks, parcel, state plan coordinates, latitude/longitude coordinates, river mile, or street address.

resolution: The resolution is that resolution at which the system is planned to work or currently works. Most of the systems involved have arbitrary resolution. That is, they may be initialized to work at any single resolution, but, once initialized, the resolution could not be changed.

coverage: Two areas of coverage are tabulated, regional and urban. Regional information systems are suitable for a single or multiple county through state or national levels. Urban systems are suitable for use in cities or metropolitan areas.

Data: Only the type and structure of the data are tabulated.

type: The type of data that may be handled is either numeric or character.

structure: The data structures used in the data base can be single attributes of a point or cell, multiple attributes of a point or cell, or tree structures attached to a point or cell.

Access: User access to the system or module is described by the mode and user interface.

mode: The access mode to a system may be batch or conversational (time-shared).

user interface: User interfaces are provided through special purpose user languages, parameter fed canned programs, or computer languages which must be used via a programmer.

Data manipulation: Only the method of creating and manipulating study regions and the type of data operators allowed are tabulated.

study regions: Study regions can be created by defining the of the study region or by defining the content of the cells or points to be included in the study region. Furthermore, some systems only allow study regions to be composed of contiguous cells. Others allow both contiguous and noncontiguous study regions to be manipulated.

operators: The data operators tabulated are the standard numerical operators (+,-,/, etc.), logical operators

(and, or, not, etc.), and some specified statistical operations (indicated in the table).

Output: Three forms of output are described: maps, report generators, and computer file generators.

maps: Maps may be produced on line printers, cathode ray tubes, pen plotters, or electrostatic plotters.

report generator: Report generators may have the capability of generating standard format tables or variable format tables under user control and possibly an ability to aggregate data into sub-classifications and distributions before printing.

file generator: File generators prepare special magnetic tape files and other machine readable files which may be sent to other software systems on the same machine or even to a different computer altogether.

Storage: The storage devices used and the efficiency of the data storage technique are tabulated.

devices: The devices most commonly used to store data on are computer punch cards, tape, and disk.

efficiency: The efficiency of storage describes how efficiently disk systems utilize space on direct access storage devices.

Low efficiency systems are defined as those systems with fixed length records and no data compression facility.

Moderate efficiency systems are those systems with variable length records and no data compression facility or with fixed length records and a data compression facility.

High efficiency systems are those systems which have variable length records and a data compression facility.

Frequently, some capabilities were planned (e.g., maps, report generators, file generators, or a user language) but no detailed capabilities were designed. In such a case, a "yes" is inserted into that portion of the table to indicate that such a facility is planned but details are not known. Some classifications and capabilities are "not applicable" to various systems and components. In a few instances capabilities were still in design phase or "insufficient data" was available to the IRIS staff at the time of table preparation to adequately describe the capability.

A single page description of each system, module, and study group follows in Table 2. These brief descriptions describe only the most important or interesting features of the system and give a brief indication of the nature of the study, the system, or the module. Individuals to contact at the various organizations and a short bibliography are included at the end of each summary.

TABLE 1

ORGANIZATIONS SURVEYED

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
Cornell University-Center for Aerial Photographic Studies	LUNR - New York State Land Use and Natural Resources Inventory	System	New York	Publications, Personal Contact
Washington - Department of Natural Resources	GRIDS - Gridded Resource Inventory Data System	System	Washington	Publications, Personal Contact
University of Massachusetts	LUIS - Land Use Information System	System	Massachusetts	Publications, Personal Contact
U. S. Department of Commerce - Economic Development Administration	CMS - Composite Mapping System	System	U.S.	Publications
Oak Ridge National Laboratory	ORRMIS - Oak Ridge Regional Moding Information System	System	U.S.	Publications, Personal Contact, Site Visit
University of Illinois - Center for Advanced Computation	NARIS - The Natural Resource Information System	System	Illinois	Publications, Personal Contact, Site Visit
University of Minnesota	MLIS - Minnesota Land Management Information System Study	System	Minnesota	Site Visit
New York City - Office of the Mayor Office of Administration	GIST - Geographic Information System	System	New York	Publications, Personal Contact

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
U. S. Department of Commerce - Bureau of the Census	DIME - Dual Independent Map Encoding	System	U.S.	Publications, Personal Contact, Site Visit
U. S. Department of Commerce - Bureau of the Census	ADMATCH	Module	U.S.	Publications, Site Visit
Canada - Dominion Bureau of	GRDSR - Geographically Referenced Data Storage and Retrieval System	System	Canada	Publications, Personal Contact
University of British Columbia	IIPS - Inter-Institutional Policy Simulator	System	Canada	Publications, Personal Contact, Site Visit
Observatire Economique Mediterranean	French Information System Network for Regional and Urban Planning	System	France	Publications
Swedish Central Board for Real Estate Data	FRIS	System	Sweden	Publications
University of Oregon Bureau of Governmental Research and Service	MAP/MODEL	System	Oregon	Publications
Canada - Department of Regional Economic Expansion	CGIS - Canada Geographic Information System	System	Canada	Publications, Site Visit
U. S. Army Corps of Engineers - Waterways Experimental Station	The Waterways System	System	U.S.	Publications, Personal Contact

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
U. S. Environmental Protection Agency - Water Quality Office	STORET	System	U.S.	Publications
Bay Area Transportation Study Commission (BATSC)	BATSC	System	California	Publications
Charlotte, North Carolina	IMIS - Integrated Municipal Information System	System	North Carolina	Publications, Site Visit
Maine - Department of Inland Fisheries and Game	MIDAS - Maine Information Display Analysis System	System	Maine	Publications, Site Visit
U. S. Central Intelligence Agency	The Auto-map System	Module	U.S.	Publications, Personal Contact, Site Visit
U. S. Army Topographic Command (TOPOCOM)	Automatic Contour Digitizer	Module	U.S.	Publications
Illinois Geological Survey	ILLIMAP	Module	Illinois	Publications, Personal Contact, Site Visit
Canadian Hydrographic Service	Canadian Hydrographic System	Module	Canada	Publications
Royal College of Art	Experimental Cartographic Unit	Module	England	Publications
Ministry of Housing and Local Government	LINMAP (Line Printer Mapping)	Module	England	Publications
Laboratory for Graphics and Spatial Analysis - Harvard University	SYMAP	Module	U.S.	Publications

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
Southern California Regional Information Study	GRIDS - Grid Related Information Display System	Module	U.S.	Publications, Site Visit
Urban Data Center University of Washington	SACS - Street Address Conversion System	Module	Washington	Publications
U. S. Department of Agriculture - Forest Service	MIADS - Map Information Assembly and Display System	Module	U.S.	Publications
Alaska - Daniel, Mann, Johnson, and Mendenhall (DMJM)	Land Use Plan	Study Group	Alaska	Publications
American Bar Association Committee on the Improvement of Land Title Records	CULDATA: A Comprehensive Unified Land Data System	Study Group	U.S.	Publications, Personal Contact
American Institute of Planners: Information Systems Department		Study Group	U.S.	Publications
Association of Bay Area Governments	Bay Region Planning Information Support Center (BRISC)	Study Group	California	Publications
Arizona State University Center for the Study of Urban Systems		Study Group	Arizona	Publications
California - Lockheed	Statewide Information System Study	Study Group	California	Publications
California - TRW	Land Use Information System Study	Study Group	California	Publications
Canadian Meteorological Service		Data Base	Canada	Publications

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
Chicago Area Transportation Study (CATS)		Study Group	Illinois	Site Visit
Denver Regional Council of Governments and Peat Marwick and Mitchell and Co.	Conceptual Design of a Regional Information System	Study Group	Colorado	Publications
Hebrew University of Jerusalem - Department of Geography		Study Group	Israel	Publications
Indiana Office of Traffic Safety	INTRAC Indiana Traffic Accident System	Data Base	Indiana	Publications
International Union (IGU) - Commission on Geographical Data Sensing and Processing		Study Group	U.S.	Publications
Los Angeles City Planning Department	LAMIS - Los Angeles Municipal Information System	Study Group	California	Publications
Metropolitan Washington Council of Governments		Study Group	District of Columbia	Publications
National Aeronautics and Space Administration-Mississippi Test Facility		Study Group	Mississippi	Personal Contact
Presidents Commission on Federal Statistics		Study Group	U.S.	Publications
Presidents Council on Environmental Quality	National Environmental Monitoring System	Study Group	U.S.	Publications, Personal Contact, Site Visit

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
Southern California Association of Governments	Regional Information System	Study Group	California	Publications
Study of Environmental Quality Information Programs in the Federal Government	SEQUIP	Study Group	U.S.	Publications
U. S. Air Force - Aeronautical Chart and Information Center		Study Group	U.S.	Publications
U. S. Department of Commerce - Bureau of the Census	New Haven Census Use Study	Study Group	U.S.	Publications
U. S. Department of the Interior - Geological Survey		Study Group	U.S.	Publications
U. S. Department of Housing and Urban Development	USAC - Federal Urban Information Systems Inter-Agency Committee	Study Group	U.S.	Publications
University of Illinois Coordinated Science Laboratory	Kankakee Project	Study Group	Illinois	Publications, Personal Contact, Site Visit
Urban and Regional Information Systems Association (URISA)	Special Interest Group in Geographic Based File Systems	Study Group	U.S.	Publications, Personal Contact
U. S. Air Force - RADC Experimental Cartographic Facility		Module	U.S.	Publications
R. L. Polk and Company	Polk Urban Information System	System		Publications

Organization Name	Project Name	Project Type	State or Country	Survey Techniques
Bureau of State Planning - Wisconsin Department of Administration		Study Group	Wisconsin	Publications

TAB

GEOGRAPHIC INFORMATION SYSTEMS AN

NAME	CLASSIFICATION		GEOGRAPHIC ATTRIBUTES			DATA		AC
	TYPE	STATUS	ACCESSING	RESOLUTION	COVERAGE	TYPE	STRUCTURE	MODE
...York State Land Use and Natural ...Inventory Center for Aerial Photo- ...Studies, Cornell, Univ.	System-uniform grid	Implemented	UTM	One square kilometer	Regional	Numeric	Multiple attributes	Batch
...Resource Inventory Data ...Dept. of Natural Resources, ...Washington	System-uniform grid	Developed	State plane coordinates	10 acre sample cells	Regional	Numeric	Multiple attributes	Batch
...Information System ...Forestry Dept.	System-uniform (dot) grid	Developed	x,y coordinate	Arbitrary	Regional	Numeric	Multiple attributes	Batch
...Mapping System U.S. Dept. of ...Economic Development Administration	System-uniform grid	Feasible	Latitude-longitude	2 x 2 mile	Regional	Numeric	Multiple attributes	Batch
...Plains Regional Modeling Informa- ...Plains National Laboratory	System-uniform grid	Experimental	Latitude-longitude	10x10 sec 15x15 sec 15/16x15/16 sec	Regional	Numeric	Trees	Batch
...Resource Information System ...Center for Advanced Compu- ...Northeast Illinois Natural ...Resource Center	System-non-uniform grid	Developed	Public rectangular land survey	40 acre 1/4-1/4 section	Regional	Numeric character	Trees	Conversational batch
...Land Information System" ...Minnesota	System-non-uniform grid	Planned	Public rectangular land survey	40 acre 1/4-1/4 section	Regional	In design	In design	In design
...Geographic Information System ...City Office of Administration ...the Mayor	System-generalized parcel (block face centroid)	Implemented	Census block, street address.	Lot-ownership parcel	Urban	Numeric character	Multiple attributes	Batch
..."Independent Map Encoding" ...Address Matching" U.S. Bureau ...	System-generalized parcel base file and address [illegible] system	Developed	Census block, census tract, etc. x,y coordinate	Census block	Urban	Numeric character	Not applicable	Batch
"Geographically Referenced Data Storage ...System" Dominion Bureau of ...Canada	System-generalized parcel (located by centroid)	Developed	UTM and street address	Block face enumeration area	Urban Regional	Numeric	Multiple attributes	Batch
...Institutional Policy Simulator" ...British Columbia	System-generalized parcel and area boundary	Feasible	x,y coordinate and address conversion	Arbitrary	Regional Urban	Numeric character	Trees and linked files	Conversational
...Information System for Regional and ...Observatoire Economique ...	System-generalized parcel	Experimental	City block	City block	Urban	Insufficient data	Insufficient data	Conversational
...Land Information System ...for Real Estate Data, Sweden	System-generalized parcel	Planned	Parcel and x,y	Ownership parcel	Regional	Insufficient data	Insufficient data	Insufficient data
...Bureau of Governmental Research and ...University of Oregon	System-area boundary	Implemented	State plane	Arbitrary	Regional	Numeric character	Any structure—but must be generated by computer programmer	Batch
...Geographic Information System" ...Department of Regional Economic ...	System-area boundary	Feasible	7 1/2 minute latitude longitude "frames"	100'	Regional	Numeric	Any structure—but must be generated by computer programmer	Batch
...Waterways Experiment Station ...Corps of Engineers	System-network	Implemented	x,y coordinate	Arbitrary	Regional	Numeric	Multiple attribute	Conversational and batch
...Quality Data Storage and ...System U.S. Environmental ...Agency	System-network	Experimental	River mile index, latitude/longitude	.1 mile	Regional	Numeric character	Insufficient data	Insufficient data
...Transportation Study	System-point	Implemented	x,y coordinate	Arbitrary	Regional	Numeric character	Trees	Batch
...United Municipal Information System ...North Carolina	System-Municipal information management system	Experimental	Insufficient data	Insufficient data	Urban	Insufficient data	Insufficient data	Insufficient data
...Information Display-Analysis ...Department of Inland Fisheries	System-tape data management system	Developed	Master block index (see comments)	Minor Civil Division	Regional	Numeric character	Multiple attributes	Batch
...Intelligence Agency	Module-cartographic	Implemented	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
...Army Topographic Command	Module-cartographic	Implemented	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
...Geological Survey	Module-cartographic	Developed	Public rectangular land survey	200 feet	Regional	Not applicable	Not applicable	Batch
...Graphic System—Canadian ...Service	Module-cartographic	Experimental	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
...Cartographic Unit—England-Pure ...of	Module-cartographic	Experimental	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
..."Printer Mapping" England- ...Housing and Local Government	Module-printer mapping program	Feasible	Insufficient data	Insufficient data	Not applicable	Not applicable	Not applicable	Batch
...ory for Computer Graphics ...sity	Module-printer mapping program	Implemented	Arbitrary x,y coordinates. No inherent coordinate system	Arbitrary	Not applicable	Numeric	Not applicable	Batch
..."...Information Display System" ...Regional Information Study	Module-uniform grid printer mapping	Developed	Arbitrary x,y coordinates, no inherent coordinate system	Arbitrary	Not applicable	Numeric	Not applicable	Batch
..."Address Conversion System" ...University of Washington	Module-street address to x,y coordinate conversion	Developed	Street address	City block	Urban	Not applicable	Not applicable	Batch
..."...Information and Display System" ...Service	Module-uniform grid base file	Implemented	Insufficient data	1/5 in. by 1/6 in. cells laid onto any base map	Single attribute	Two digit character codes	Insufficient data	Batch

PREPARED FOR THE
FOR ENVIRONMENTAL
CENTER FOR ADVANC
UNIVERSITY OF ILLI

RELATED COMPONENTS—JANUARY 1, 1972

	DATA MANIPULATION		OUTPUT			STORAGE		
ER INTERFACE	STUDY REGIONS	OPERATORS	MAPS	REPORT GENERATOR	FILE GENERATOR	DEVICES	EFFICIENCY	COMMENTS
r language	Defined by coordinates	Numeric logical	Line printer	Standard format tables	No	Disk	Low-fixed length records no data compression encoding	Only working statewide system
ed programs meters)	Defined by coordinates	Numeric (totals only)	Line printer	Standard tables	No	Disk	Insufficient data	Based upon MIADS work.
ed programs meters)	Not applicable	Numeric	Line printer	Standard format tables	No	Cards	Not applicable	This is a recent (1971) set of canned programs, being used as a pilot for a statewide system. Currently the system is limited to 3000 data (grids).
ed programs meter)	Defined by coordinates	Numeric	Line printer	Not applicable	Not applicable	Tape	Not applicable	Designed for very large region, low resolution printer mapping.
outines called FORTRAN	Defined by coordinates	Numeric	Printer	Not available	No	Disk	In design	Being designed to support specific regional models.
language	Defined by content/ coordinates, need not be contiguous	Numeric logical	Line printer and electrostatic plotter	Variable format tables aggregation into sub classifications	Yes	Disk	High-variable length records with data compression encoding	System has been frequently demonstrated. Scheduled to enter full time service July, '72. Only non-experimental geographic information system that is completed.
esign	In design	In design	Line printer	In design	In design	In design	In design	Mapping programs are in an experimental state. Rest of the system is undefined.
ed programs ameters)	Not applicable	Not applicable	Line printer	Not applicable	Not applicable	Tape	Not applicable	Designed to handle data on buildings, property and streets. GIST is ADMATCH with a New York Address Coding Guide and SYMAP with a New York base file for mapping.
ed programs ameters)	Not applicable	Not applicable	Pen plotter	Standard format table	Yes	Tape	Not applicable	The DIME system is a way of organizing urban census data. ADMATCH is a program for converting street addresses to census tracts.
u uage	Defined by block faces	Numeric (totals averages)	Line printer	Fixed format	Yes	Disk	High	GRDSR allows a user to tabulate census statistics on an area he selects by describing the block faces which enclose it.
u language	Defined by Coordinates content	Numeric logical	CRT penplotter electrostatic	Uses Cornell-ASAP	Planned	Disk	Moderate-variable length rec.	Highly graphic oriented designed primarily to integrate with simulation package
ed programs meters)	Insufficient data	Insufficient data	Yes	Yes	Insufficient data	Insufficient data	Insufficient data	A non-sophisticated, limited, but reasonable project to handle census data.
fficient data	Insufficient data	Insufficient data	Yes	Yes	No	Tape	Not applicable	This is an ambitious parcel system project for Sweden. An augmented computer language is considered for use as a user interface language. The whole system is in the design phase
programmer 1 language)	Defined by Coordinates	Calculates: areas length location of centroid	Pen plotter	No	No	Tape	Not applicable	Better algorithms than CGIS however, it requires a PL/1 programmer to use this system.
programmer language)	Not applicable	Not applicable	Planned	No	No	Tape	Moderate-variable length records no data compression encoding	Very expensive data input and output. Requires a computer progr. to use it and excessive amts. of core storage. Should possibly be classified as a base file module rather than a system
programmer FRAN language)	Not applicable	Not applicable	Pen plotter	Requires programming	Not applicable	Tape	Not applicable	Limited information available. Special purpose system to handle waterways problems
language (to ulate canned ms)	No	Insufficient data	Pen plotter	Standard format tables	No	Disk	Insufficient data	Contracts have been let to digitize the stream network of half the country. This will cover all but one basin in Illinois.
ed programs ameters)	Defined by coordinates	Numeric logical statistics	Pen plotter	Standard format tables	Yes	Tape	Not applicable	The system is made up of three System Development Corporation Systems, SPAN, NALAP, DATADOX.
fficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data	Insufficient data	This system was planned to have a geo-coding capability. Other serious problems have prevented attending to the geographic nature of their data.
report erator language	Not applicable	Numeric	Very limited line printer	Variable format tables, aggregation into subclassifications	No	Tape	Not applicable	Conventional, well done tape data mgt. sys. 200 byte master block index contain's co., twp. UTM coordinates, etc. added to each record. Sec refg sis sort on master block index fields.
applicable	Not applicable	Not applicable	Pen plotter	Not applicable	Not applicable	Not applicable	Not applicable	Very nice digitizer program with multiple projection capability.
plicable	Not applicable	Not applicable	Pen plotter	Not applicable	Not applicable	Not applicable	Not applicable	This is an automated digitizing program
ed programs ameters)	Not applicable	Not applicable	Pen plotter	Not applicable	Not applicable	Not applicable	Not applicable	ILLIMAP will draw SINGLE SITES e.g. well location, political boundaries and sections given a legal description of the area or point in Illinois.
plicable	Not applicable	Not applicable	High accuracy pen plotter	Not applicable	Not applicable	Tape	Not applicable	This cartographic system includes an interactive CRT to aid digitizing and map presentation efforts.
plicable	Not applicable	Not applicable	Pen plotter	Not applicable	Not applicable	Tape	Not applicable	System is being designed to provide on-line interaction and editing directly linked to a digitizer
ed program ameters)	Defined by coordinates	Not applicable	Line printer	Not applicable	Not applicable	Insufficient data	Not applicable	Similar to SYMAP module—the production of color maps is beginning to be researched.
ed program ameters)	Defined by coordinates	Not applicable	Line printer	Not applicable	Not applicable	Cards	Not applicable	Heavily used as a mapping module in many systems. University of North Carolina planned up and completed the initial work at Harvard.
language al)	Defined by coordinates	Not applicable	Line printer	Not applicable	Not applicable	Cards, tapes	Not applicable	Coded in ASA standard FORTRAN IV for exportability to other systems. Cells 1 to 99 characters on a side may be printed
plicable	Not applicable	Not applicable	Pen plotter	Not applicable	Not applicable	Tape	Not applicable	This system is similar to the Census Bureau ADMATCH but was designed for small computers and geographic accuracy.
ed programs ameters)	Not applicable	Not applicable	Line printer	No	No	Punch cards	Not applicable	This is a limited capability system for storing two digit codes on uniform grid cells. MIADS and [illegible]MAP are follow-on systems

LLINOIS INSTITUTE
RITY BY THE
D COMPUTATION,

Organization: Center for Aerial Photographic Studies, Cornell University

System: New York State Land Use Natural Resources Inventory (LUNR)

Comments: LUNR is the first complete statewide information system to possess any data input checking and error correction as well as retrieval and graphic capabilities. Data is entered by cards onto disks. The system provides no data compression facilities. Much of the disk space is unused. A user must go to a code book to interpret the data.

Retrieval and analysis of the data are facilitated by two computer programs, DATALIST for tabular output, and PLANMAP for graphic output. All data in the inventory system are stored on IBM 2316 disks with random access storage and retrieval by UTM GRID cell coordinate numbers.

A card-oriented batch processing program, a user-oriented language, and report generation facility are provided.

Funding Agency: New York State Office of Planning and Coordination

Contact: D. Belcher, Director of Center of Aerial Photographic Studies, Cornell University, Ithaca, N. Y.

References: D. Belcher, et. al. New York State Land Use Natural Resources Inventory Final Report. Vol. 1-5, Center for Aerial Photographic Studies, Cornell University, Ithaca, New York, 1971

Organization: Department of Natural Resources, State of Washington

System: Gridded Resource Inventory Data System (GRIDS)

Comments: The GRIDS system appears to be an advancement of the MIDAS system. Data is collected once on acre samples within a ten acre parcel. The data is referenced by section coordinates of the legal land survey. Map printout by township are printed and each print position represents a ten-acre grid point (1/4 1/4 1/4 section).

The grid sampling locations extend over all the land administered by the Department of Natural Resources. The sampling locations are spaced at approximately 660 foot (ten chain) on the Washington State plane coordinate system. Base maps are generated from topographic maps and aerial photographs at a scale of 1 inch = 1000 feet to identify the sampling locations. The system is capable of storing multiple attributes of each sample point in a single record.

Funding Agency: The Department of Natural Resources, State of Washington

Contact: Roger A Harding, Inventory Supervisor, Technical Services Division, Department of Natural Resources, State of Washington, Box 168, Olympia, Washington, 98501

References: R. A. Harding, "Resource Inventory by GRIDS." Technical Services Division, Department of Natural Resources, Box 160, Olympia, Washington 98501, April, 1969.

R. A. Harding, DNR Technical Services Manual, 010205 GRIDS - A Manual for Data Entry, Department of Natural Resources, Box 160, Olympia, Washington, 98501, April 1969.

"GRIDS - File Building and Editing," internal document, Department of Natural Resources, Box 160, Olympia, Washington, 98501, April 1969.

Organization: University of Massachusetts, Department of Forestry

System: Land Use Information System (LUIS)

Comments: The LUIS system consists of a package of four FORTRAN IV computer programs which are designed to accept coded information for a series of data points associated with a given source map. A conventional dot grid provide information points to which items of a map can be attached. Some of its most useful features are that:

a.) It can compute the area and perimeter of all discrete land units and total them in various ways.
b.) It can compute water frontage for lakes, rivers, and streams.
c.) It can compute road frontage
d.) It can retrieve discrete land units with various restrictions as to size, location, or as juxtaposition to roads, buildings, etc.
e.) The retrieved land units can be displayed in a computer map form similar to SYMAP.
f.) Changes in land use between two or more time periods can be analyzed and the results portrayed in a tabular or map form.

The depths to which these analyses can be formed is not clear because of a lack of documentation.

Funding Agency: Federal Government under the MacIntre-Stennis Bill.

Contact: William P. MacConnel, Forestry Department, University of Massachusetts, Amherst, Massachusetts. Richard A. Howard, Director of Computer Center, Clark University, Worchester, Massachusetts.

References: MacConnel, W.P., and Howard, R.A., LUIS-Map Information Manipulation, Retrieval, and Display System," ACSM March 7, 1971, p. 340.

Organization: Office of Planning and Program Support Economic Development Administration Commerce Department

System: Composite Mapping System (CMS)

Comments: The Composite Mapping System (CMS) is a uniform grid system designed primarily for large grids (2 x 2 miles) covering vast areas and to be used to analyze census data. CMS covers the country in 240 x 240 mile sectors. Thus, each sector covers 4 degrees latitude by 4 degrees longitude and each grid covers 2 minutes by 2 minutes. While most systems store all the different data attribute values for a single grid in one record, the CMS will store one attributes' value for all the grids in a sector as a single record. Therefore, the spatial printing of one record in CMS will be a map of one data attribute. The system allows several maps to be stored that cover the same sector thus allowing several attributes to be stored for each sector and grid. By permitting only numerical data values, the system allows several maps to be overlapped or composited by considering each map or data attribute over a sector to be considered as a matrix (i.e., an array of spatial related data values) and that several of these matrices can be weighted and added together. The final composite map is output to the user on a line printer who performs any spatial analysis by visually scanning the output.

Funding Agency: U.S. Department of Commerce

References: Nez, George, "A Composite Mapping System for Practical location Research", Office of Planning and Program Support, Economic Development Administration, Commerce Department.

Organization: Oak Ridge National Laboratory

System: Oak Ridge Regional Modeling Information System (ORRMIS)

Comments: ORRMIS is a uniform grid system having a variable size grid structure to support the Oak Ridge regional modeling efforts in intrinsic land use suitability, land use simulation, soil loss prediction, and social economic modeling. The system is batch oriented and supports specific, modeling activities. It incorporates sophisticated scanning hardware to read maps and photographs, plus hand-supplied overlay information and encodes this into the various grid structures available in the ORRMIS system. The most popular cell size is 140 acres (30 x 30 secs.) although other cell sizes are available, from 15 sec. on a side to 15/16 sec. on a side. The 15/16 sec. cell is being used only for topographic information at this time.

Funding Agency: National Science Foundation. Interdisciplinary Research Relevant to the Problems of Our Society. (IRRPOS)

Contact: Jack Gibbons, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

Organization: University of Illinois, Center for Advanced Computation, in cooperation with the Northeast Illinois Natural Resource Service Center Center

System: Natural Resource Information System (NARIS)

Comments: This system which stores attributes of 40 acre 1/4-1/4 sections, has an addressing scheme based upon the public rectangular land survey system. The system has sophisticated structuring facilities and allows one to store data on cell attributes in terms of basic classes such as soil, geology and hydrology and basic attributes of each of those classes, such as soil slopes and soil types. Many classes are allowed but only those classes present in the cell are recorded. Internally, the data are stored in a compressed format.

Study areas can be defined in terms of their coordinate locations or in terms of the cell attributes. Such regions of study need not be contiguous. This is the only complete geographic information system that allows conversational or exploratory access to the data base. A specially designed user interface language allows a resource planner to ask detailed and complex questions of the data base in a real time environment.

Output is in the form of tabular data or map data. An addition is being developed to produce low resolution. teletype and line printer maps as well as high resolution, electrostatic plotter maps.

Funding Agency: Ford Foundation

Contact: [redacted] Center for Advanced Computation, University of Illinois, Urbana, Illinois 61801

References: "NARIS: A Natural Resource Information System." Center for Advanced Computation, University of Illinois.

W. McTeer, et.al., "NARIS User's Manual" documentation. Center for Advanced Computation, University of Illinois, Feb. 1971.

Organization: University of Minnesota

System: Minnesota Land Information System (MLIS)

Comments: MLIS, a system being planned by the University of Minnesota for statewide use, was inspired by the statewide land use survey conducted by the University Center for Urban and Regional Planning and the Minnesota State Planning Agency.

It is proposed that the data base collection unit be a 1/4-1/4 section of the legal land survey. Although no system has been built, some mapping programs exist (Coded in FORTRAN for the CDC-6600) which do not account for irregularities that occur due to survey. For this reason, computer output is corrected by hand before photographing for final use.

Funding Agency: Minnesota State Planning Agency

Contact: George Orning, Department of Geography, University of Minnesota, St. Paul, Mennesota 55101

Organization: New York City, Office of Administration and Office of the Mayor

System: Geographic Information System (GIST)

Comments: The System was designed to coordinate the exchange of information concerning public and private property throughout the city government. GIST maintains files of general interest data concerning the buildings, property, and streets and sidewalks of New York City. The system contains three major files: the street name file which contains the official name and commonly used names for every street, the block face file which contains one record per block side, and the building lot file which contains one record per lot, major structure and owner. The system is actually two programs. A program similar to ADMATCH is used to do address matching, thus determining to which census tract or census block a particular address belongs. The system incorporates Harvard's SYMAP mapping program and provides users with a New York base file for mapping.

Funding Agency: City of New York, Office of the Mayor

Contact: Robert Amsterdam, Office of Administration, Office of the Mayor, 250 Broadway, New York City, New York

References:

R. M. Amsterdam, "GIST, A Geographic Information System for New York City (preliminary design)." Office of the Mayor, New York City, July, 1968.

R. M. Amsterdam, "Development of New York City's Geographic Data Network." Spring Joint Computer Conference, 1969.

"GIST, New York City's Geographic Information System - Initial Operating Capabilities and Future Development." Office of the Mayor, New York City, September, 1970.

R. M. Amsterdam, An Introduction to GIST New York City's Geographic Information System, Office of the Mayor, May 1971.

Organization: U. S. Department of Commerce, Bureau of the Census

Module: The DIME Geo-coding System and ADMATCH Program

Comments: The DIME (dual independent map and coding) system was developed for the use of the Census Bureau. DIME is used to code data on areas and boundaries (usually streets that enclose those areas). DIME is normally used with the ADMATCH system to match street addresses with blocks. ADMATCH is a computer program written for the New Haven Census Use Study.

DIME files contain a representation of the graphic layout of the street network. They allow data to be stored on the areas enclosed by the street network which are easily related to other parcels of data or other blocks.

The ADMATCH program, with an associated ACG (address coding guide), can be used to match street names and addresses with census blocks. However, the general success of this program has been poor. A replacement for ADMATCH is being generated at the Bureau of the Census. It is based upon classification theory, and should be a more useful piece of computer software.

Funding Agency: U. S. Department of Commerce, Bureau of the Census

Contact: Jim Corbett, U. S. Department of Commerce, Bureau of the Census, Washington, D. C.

References:

"Census Use Study Report #2, Computer Mapping." U.S. Department of Commerce, Bureau of the Census.

"Census Use Study Report #4, The DIME Geo-coding System." U.S. Department of Commerce, Bureau of the Census.

"Census Use Study, ADMATCH Users Manual." U.S. Department of Commerce, Bureau of the Census.

"Geographic Supervisors Manual." U.S. Department of Commerce, Bureau of the Census. Geo 70-108, May, 1960.

"Geographic DIME File Coders Manual." U.S. Department of Commerce, Bureau of the Census. Geo 70-108, May, 1960.

Organization: Dominion Bureau of Statistics, Canada

System: Geographically Referenced Data Storage and Retrieval System (GRDSR)

Comments: The Dominion Bureau of Statistics has developed a computerized system for providing census data for the 1971 census on a user-specified basis. By describing the block faces which enclose it, a study area can be specified and statistical summaries on that area can be requested. Summaries are restricted by confidentiality requirements of the Statistics Act.

The main objective is to provide tabulations relatively quickly and inexpensively by automatic selection and by aggregations of a series of building blocks that make up the user-specified area. In large urban areas these are city block faces; elsewhere they are enumeration areas.

A DIME-like system is the heart of GRDSR. Street patterns are digitized and given geographic coordinates in the UTM System. An address-matching capability allows all addresses to be automatically mapped into UTM coordinates and then into census blocks and enumeration areas.

Data is stored on 22 disks using a special coding structure that allows 60% reduction in the total amount of storage space formerly required. A special user language is also provided to prepare the reports on user-specified study areas.

Funding Agency: Dominion Bureau of Statistics, Canada

Contact: John Welson, Chief of General Survey Systems, Sampling and Research Staff, Dominion Bureau of Statistics, Ottawa 3, Ontario

References:

"Geo-coding facts by small areas." Bulletin #1 Dominion Bureau of Statistics, Ottawa 3, Ontario. February 1969.

R. F. Tomlinson, Editor, "Environmental Information Systems" Proceedings of the UNESCO/IGU First Symposium on Geographical Information Systems. Ottawa, Ontario. September 1970. pp. 76, 77, 123 and 124.

"Statistic Canada--Introduction to the GRDSR System," Presented at the Provincial Census Data Access Workshop, Ottawa, October 17, 1971.

Ion, R. J., et. al., "GRDSR Geographical Referenced Data Storage and Retrieval System," Dominion Bureau of Statistics, Ottawa, 1970.

Organization: University of British Columbia

System: Inter-Institutional Policy Simulator (IIPS)

Comments: The IIPS is being coded at the University of British Columbia with equal cooperation from the University, the city of Vancouver, and the Vancouver Regional Board. The purpose of the project is to simulate the effects of policy decisions for the region. Output will include several display devices which will present maps and graphs. In addition, tabular output from the simulators will be available.

There are four integrated subsystems in the IIPS: a supervisor, a graphic supervisor, an information supervisor, and the simulation packages. In addition, a command language and data language will be provided for the system.

Funding Agency: Ford Foundation

Contact: C. S. Hollings, University of British Columbia, Vancouver, British Columbia, Canada, or J. L. Parker, University of British Columbia, Vancouver, British Columbia, Canada

References: James L. Parker, "A Graphics and Information Retrieval Supervisor for Simulators." Department of Computer Science, University of British Columbia. July 28, 1971.

James L. Parker, "The Natural Information Systems Project, An Overview." Department of Computer Science, University of British Columbia. August 8, 1971.

James L. Parker, "Information Retrieval with Large Scale Geographic Data Bases." Department of Computer Science, University of British Columbia. Report #1, June, 1971.

Organization: Observatoire Economique Mediterranean

System: French Information System Network for Regional and Urban Planning

Comments: The OEM Geographic Information System is an experimental time-shared and batch information system. Data is stored on the last two censuses in France and is aggregated on the block level for each city with more than 20,000 people in the south of France. Data is also available on aggregate units of a single rural city or a single urban district.

Retrieval is done via "Macro Instructions" which provide a capability to list or sort multi-criteria data. Output is available at a remote terminal via a CRT screen or a line printer. Although computer mapping is available, it is not clear whether mapping is via line printer or CRT.

The OEM people depict their system as a "non-sophisticated, limited, but reasonable project."

Contact: Gean Salmona, Observatory Economic Mediterran, Marseille, FRANCE

References: Tomlinson, R. F., "Environmental Information Systems," _Proceedings of the UNESCO/IGU, First Symposium on Geographic Information Systems_, Ottawa, Canada, October 1970.

Organization: Swedish Central Board for Real Estate Data

System: A Spatial Information System (FRIS)

Comments: FRIS is a generalized parcel system which is being planned for implementation across the whole of Sweden. The pilot system is being done on a 530 square mile area.

The basic unit of data in the FRIS system is the ownership parcel. The system can record data on x,y coordinates which are part of a special coordinate for parcel data, all data recorded within a parcel are assigned to an approximate centroid point for that parcel.

While the system is planned to be regional or country-wide, the use of a parcel level coding system seems to be more appropriate to an urban system. This, along with the small size of the initial study areas, raises significant doubts of the applicability of the system over the whole country.

Contact: Ove Salmonsson, Swedish Central Board of Real Estate Data, Sundbyberg, Sweden

References:

Tomlinson, R. F., "Environmental Information Systems." The Proceedings of the UNESCO/IGU, First Symposium on Geographical Information Systems. Ottawa, Canada, October 1970.

Alfredsson, B., et.al., "A Spatial Information System Introduction." FRIS A:1, 1970, Sundbyberg, Sweden.

Selander, Kay, "A Spatial Information System, Registration and Storing of Coordinates," FRIS C:1. October, 1970.

Olsson, A. and Selander, Kay, "A Spatial Information System Dot Map by Computer." FRIS C:2. May, 1971.

Olsson, A., "A Spatial Information System Program for Coordinate Processing." FRIS C:3. May, 1971.

Organization: Bureau of Governmental Research and Service, University of Oregon

System: MAP/MODEL

Comments: This system, like the Canadian Geographic Information System, is a map image processing system; however, its algorithms are more efficient. The user is provided with boundary editing, measurements of area length and centroid location, and plotting routines for point, line, and polygon data. Alpha-numeric retrieval, sorting, and summarization are provided through a PL/1 language interface and therefore, require programmer intervention. Future plans include report generation and some on-line capability.

Funding Agency: The Bureau of Governmental Research and Service

Contact: Robert E. Keith, School of Community Service and Public Affairs, University of Oregon, Eugene, Oregon 97403

References:

Arms, S., "MAP/MODEL System-Technical Concepts and Program Description." Columbia Region Association of Governments, Portland, Oregon.

Arms, S., "MAP/MODEL System - System Description and Users Guide." Bureau of Governmental Research and Service, University of Oregon, May, 1970.

Organization: Canadian Department of Regional Economic Expansion

System: Canadian Geographic Information System (CGIS)

Comments: CGIS is being developed by the Department of Regional Economic Expansion to support the Canadian Land Inventory. The Canadian Land Inventory collects data on land use activity suitabilities and interpretations pertaining to agriculture, forestry, recreation, wildlife, etc., and present land use classifications. Most suitabilities are scored on a scale of 1 to 9.

CGIS, an area boundary system which provides map overlays as its main capability, was developed by the IBM Corporation in Canada. This system has a basic resolution capability of approximately 100 feet. No user interface language is supplied. All requests must be interpreted by a programmer and then written in the IBM computer language PL/1. A PL/1 program thus written will overlay the maps, prepare appropriate tables as desired, and provide graphic output.

To date, the cost of the data collection effort has exceeded 20 million dollars. The cost of the computer system development has exceeded 3 million dollars. Unfortunately, the operational costs of the computer system are so high that use on the provincial level is impractical. Thus, it is relegated to experimental use at the federal level.

Funding Agency: The Department of Regional Economic Expansion under Agriculture Rehabilitation and Development Act Funds

Contact: David Glover, Information Systems Division, Department of Regional Economic Expansion, 473 Albert Street, Ottawa, Ontario, Canada

References:

J. B. McClellan, "The Land Use Sector of the Canada Land Inventory," Geographic Bulletin. 1965. Vol. 7, No. 2, pp. 73-78.

Tomlinson, "An Introduction to the GEO-Information System of the Canada Land Inventory." Canada Department of Forestry and Rural Development. Ottawa, 1967.

Organization: U. S. Army Corps of Engineers, Waterways Experiment Station

System: The Waterways System

Comments: The Waterways System is a point geographic system that stores gauge reading with respect to time, channel slopes, soils and other hydrographic data. The system will plot hydrographs of water channel discharges. A bridge emplacement model is coupled with retrieval capabilities. Also, a flow model is incorporated to model problems of sewage collection by being able to predict the flow of effluents at any point in the channel network.

This system was intended to support the Army in tactical situations - hence, the emphasis on bridge emplacement.

Funding Agency: U. S. Army Combat Developments Command

References:

Friesz, R. R., et. al. "A European Waterways Study--A Procedure for Describing Tactical Gaps", Volumes 1 and 2, Technical Report M-70-12, U.S. Army Waterways Experiment Station, Vicksburg, Mississippi, July 1970.

Organization: U. S. Environmental Protection Agency

System: Water Quality Data Storage and Retrieval System (STORET)

Comment: Work on STORET concepts was initiated prior to 1966 in the Department of Public Health. STORET can store and retrieve stream network and point effluent and sampling data. These capabilities are needed for water quality monitoring and enforcement programs. The STORET system will involve digitizing all the stream networks in the nation. Bids have been recieved and contracts awarded for the digitizing of the first half of the nation. This digitizing effort will include all of Illinois.

The latitude and longitude of point data and the stream network are digitized. The digitized points along the network are used to compute the basic address for water quality data--a river-mile index. A number of papers have been prepared on locators and the development of locator data for the STORET system.

Funding Agency: U. S. Environmental Protection Agency

Contact: Phillip L. Taylor, Room 232, Building 2, Crystal Mall, 1921 Jefferson Davis Highway, Arlington, Va. 22202

References: Green, R. S., et. al., "Data Handling Systems and Water Pollution Control," Sanitary Engineering Division Proceedings of the American Society of Civil Engineers, February 1966. p. 55.

"Location Coding for the STORET System" (Revised November 1968). U. S. Department of the Interior, Federal Water Pollution Control Administration, Division of Technical Support, Pollution Surveillance.

"AUTOMAP--A System that Determines River Mileages and Latitude/Longitude for Input to the EPA/OWP/s STORET System for Retrieval of STORET Information in Hydrological Order," Environmental Protection Agency, Office of Water Programs, Washington, D. C.

Taylor, P., "STORET--A Data Handling System in Water Pollution Control," ASCE Annual Meeting, Sanitary Engineering Division, October 13-17, 1969.

Organization: Bay Area Transportation Study Commission (BATSC)

System: BATSC Information System

Comments: BATSC is charged with planning long-range regional, ground transportation for the nine-county San Francisco region. BATSC was established by the California State Legislature to conduct a comprehensive transportation study and to prepare a master regional transportation plan for the San Francisco region. The information processing system employed at BATSC was intended to provide a greater data base capability than that available to similar studies prior to 1963. System Development Corporation was contracted to provide the software for handling the study. An extensive and growing data base, physically stored on 1100 reels of magnetic tape, consists of data acquired from field surveys and from public, federal, state, local, and private agencies. BATSC works closely with the Association of Bay Area Governments, the area's regional planning agency. These software packages are provided by the System Development Corporation: SPAN, MADAM, and DATADOX. The MADAM System was used while the SPAN System was being upgraded from the Penn-Jersey Transportation Study in late 1962. This was the first attempt to provide a data base management and statistical analysis package for socio-economic, ecological, and other statistical data used by non programmers. The DATADOX System catalogues the several thousand variables in files associated with this large tape data management system.

Funding Agency: U. S. Department of Housing and Urban Development

References: Keveny, Michael J., "An Information System for Urban Transportation Planning, the BATSC Approach." Technical Memorandum #3020/000/01. System Development Corporation, Santa Monica, California.

Organization: City of Charlotte, N. C.

System: Integrated Municipal Information System (IMIS)

Comments: IMIS is a USAC-initiated program. (The Urban Information Systems Interagency Committee (USAC) is headed by HUD.) A grant of three million dollars was received by the City of Charlotte; of this, two million dollars was spent on data systems by the System Development Corporation (SDC). The University of North Carolina received approximately $260,000 to monitor system development and evaluate its impact on government efficiency and policy making.

According to University of North Carolina (UNC) monitors the experiment seems to have failed. At UNC the general feeling is that program failure is due, at least in part, to USAC's placing emphasis on computing systems rather than the problems of implementing such systems in government. USAC officials see city hall made efficient with on-line CRT's and terminals, but no social or procedural environment presently exists for such systems in municipal governments.

In addition to these reservations, software development and hardware capabilities were questioned. The city was to contribute programming services as part of their share of the matching funds but, unfortunately, their programming talents are limited. Although SDC would like to do the programming, they claim they have no funds. RCA offered Charlotte a package deal on hardware, proposing to use Charlotte as a test site for their IPS 70 System. This was an untested system and the little software that exists for it seems to perform poorly for information management.

Geo-coding is not important to Charlotte at this time. While the city hopes to make the address coding guide/DIME system operational in the future, too many problems are more pressing at the present.

UNC does not foresee development of any prototype system in Charlotte. Problems there are too specific to the area and relationships between hardware, software, and government functions are too tight to allow for generalization.

Funding Agency: HUD Urban Information Systems Interagency Committee

Contact: John M. Plante, Technical Director, Charlotte IMIS Project, System Development Corporation, Charlotte, N.C.

Reference: "Federal Agencies Sponsor Prototype Information System." Urban Information Systems Report, Vol. 1, No. 1. September, 1970. pp. 1-2.

City of Charlotte Integrated Municipal Information System Orientation and Briefing Guide." Charlotte Municipal Information System Project, Charlotte, N. C., 15 July, 1970. Available from National Technical Information Service, PB196458.

J. M. Plante, "Status Report on USAC Municipal Information Systems Project in Charlotte." 8th URISA, 1970, pp. 55-62.

Organization: Maine Department of Inland Fisheries and Gaming

System: Maine Information Display-Analysis System (MIDAS)

Comments: The MIDAS system is being built for the State of Maine by North American Rockwell. This system is a tape data management system built on RCA computers. Each record in the system is augmented with a 200-byte master block index field which contains country, township, UTM codes, and other geographic locators. The system contains a good report generator facility and one can sort on the fields containing the master block index to aggregate data according to some geographic locator. For example, one can sort on the county field to produce summary reports of county data. In general, UTM coordinate fields represent the centroid of the area involved. The lowest resolution of the system is normally the minor civil division or township.

This system is augmented with a BIOMED statistical system (from UCLA) for data analysis and reporting. The system is more a general purpose information management system than an environmental or geographic information system. It is, however, quite adequate for use in the State of Maine because of the nature of Maine's relatively simple economic, natural resource, and demographic interactions.

Funding Agency: State of Maine Department of Inland Fisheries and Game

Contact: Bob Young, Department of Inland Fisheries and Game, State Office Building, Augusta, Maine

Organization: Central Intelligence Agency

Module: The AUTOMAP System, CAM-Cartographic Automatic Mapping Program

Comments: The Automap System is a fairly sophisticated digitizer input system. AUTOMAP can read maps in any one of nine different projections and can store them in an internal data format. Maps can then be drawn from this internal data file in any one of nine projections.

The primary input program BULAP handles inputs from maps drawn using Lambert coordinates, Universal Transverse Mercator coordinates, or polar stereographic projections. These programs run on an IBM System 360, Model 65 or Model 67.

The AUTOMAP System includes a data tape which has 65,000 points of coastline data and 30,000 points of political boundary data which may be used to plot any country in the world.

Documentation and data tapes for this system are available from the National Technical Information Service, Springfield, Va. 22157. The documentation is available via accession number, PB19779, for $3.00. The data tape containing the coastline and political boundary data and the AUTOMAP programs are available via accession number, PB19780 and cost $55.00.

Funding Agency: Central Intelligence Agency

Contact: Warren F. Schmidt, Central Intelligence Agency, McLean, Va.

References:

"Cartographic Automatic Mapping Program (CAM) Documentation," PB191779 Central Intelligence Agency, Washington, D. C., May 4, 1970.

Schmidt, Warren F., "The AUTOMAP System," Surveying and Mapping, Vol. 29, No. 1, March 1969, p. 101.

Schmidt, Warren F., "Automation and Thematic Cartography", American Congress on Surveying and Mapping, March 1-6, 1970. p. 217.

Tomlinson, R. F., Editor, Environmental Information Systems, Ontario, Sept. 1970.

Organization:	U. S. Army Topographic Command
Module:	Automatic Contour Digitizer
Comments:	This program has developed software to significantly automate the map preparation and digitizing process for data entry.
Funding Agency:	United States Army
Contact:	Wesley H. Shepherd, Project Engineer, Advanced Mapping Division, Department of the Army, the Engineer GEODESY Intelligence and Mapping Research and Development agency, Fort Belvoir, Va. 22060.
References:	Min, P. J., and Thompson, D. R., "Computer Aided Mapping, A Total Systems Approach." ASP. March 7-12, 1971.
	Shepherd, W. H., Automatic Contour Digitizer. Paper presented at the March 1967 ASP/ACSM Convention, Washington, D. C.

Organization:	Illinois Geological Survey
Module:	ILLIMAP
Comments:	ILLIMAP is a mapping program developed and implemented by the Illinois Geological Survey to allow the printing of 7 1/2 minute maps and 15 minute maps of survey sections in Illinois and specific site locations such as well locations. The system is coded in FORTRAN and has completely digitized all the section corners in the State of Illinois. Areas are described in terms of the rectangular survey coordinates.
Funding Agency:	Illinois Geological Survey
Contact:	John Frye, Chief, Illinois Geological Survey, University of Illinois, Urbana, Illinois 61801

Organization: Canadian Hydrographic Service

Module: The Canadian Hydrographic System

Comments: The Canadian Hydrographic System is a cartographic module of a geographic information system. It is designed to produce very high accuracy maps from engineering drawings of the hydrographic survey data available in the Canadian Hydrographic Service. Various components of the system run on a PDP-8 and an IBM 360 computer. A digitizer with a resolution of .0004 inches and a Gerber 32 plotter with a resolution of .0001 inches are used for high accuracy map production.

The system allows both map digitizing and map generation and provides programs that allow a user to use an interactive CRT display to aid the digitizing and map presentation efforts.

Funding Agency: Committee on Geographic Service

Contact: Ray Boyle, University of Saskatchewan, Saskatoon, Canada

References:

Boyle, A. R., "Automation and Hydrographic Charting", The Canadian Surveyor, Vol. 24, No. 5. December, 1970, pp. 519-537.

Tomlinson, R. F., "Environment Information Systems", Proceedings of the UNESCO/IGU First Symposium on the Geographical Information Systems, Ottawa, Canada, October, 1970.

Organization: Royal College of Art, London

Module: Experimental Cartographic Unit

Comments: This module is being designed to incorporate a computer on-line with a digitizer to provide interaction and editing, directly relating to digitizing. This is similar to the work being done by the Canadian Hydrographic System.

Funding Agency: National Environment Research Council, U.K.

References:

Evans, I. S., The Implementation of an Automated Cartography System.

Tomlinson, R. F., "Environmental Information Systems," Proceedings of the UNESCO/IGU First Symposium on Geographic Systems, Ottawa, October, 1970.

Organization: Ministry of Housing and Local Government, England

Module: Line Printer Mapping (LINMAP)

Comments: The capabilities of the system are very similar to those of of SYMAP. LINMAP can also print statistical histograms. The production and use of color maps is being investigated.

References: Gaits, G. M., "Thematic Mapping by Computer," The Cartographic Journal, June 1969.

Organization: Laboratory for Computer Graphics and Spatial Analysis, Harvard University

Module: SYMAP

Comments: SYMAP is a very popular line printer mapping program which has capabilities that allow it to plot points, lines and polygons as well as shade the interior of polygons. In addition, SYMAP has a number of statistical support options which permit the calculation of means, deviations, histograms, and percentile groups within the same mapping program package.

The Census Use Study Report #2 on computer mapping, although generally pleased with the level of development of the system, made the following criticisms:

> Within the SYMAP System changing from one type of map, using one mapping option, to another type of map is not easy. For instance, switching from a map using irregular polygon areas, as in conformal block shading, to a map using single data points as in contour shading requires the input of a completely rearranged geographic base. In addition, the visual attractiveness of maps produced by SYMAP was sometimes disappointing and local users tried to add cosmetics by hand to make them more presentable as well as facilitate orientation.

Funding Agency: Ford Foundation

Contact: Carl Steinitz, Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Massachusetts 02138

References: "Census Use Study, Report #2, Computer Mapping." U.S. Department of Commerce, Bureau of the Census.

Maxfield, William H., "Computer Mapping of Census Aggregated Data, the New Haven Census Use Study Experience." Proceedings of the Sixth Conference, 1968, p. 184.

"Users Reference Manual for Synagraphic Computer Mapping SYMAP Version 5", Laboratory for Computer Graphics and Special Analysis, Harvard University, Cambridge, Massachusetts. 1968.

"Programming Guide to SYMAP Version 4.5", Department of City and Regional Planning, University of North Carolina, Chapel Hill, N. C., October 1968.

Organization: U. S. Department of Commerce, Bureau of the Census, in cooperation with the Southern California Regional Information Study.

Module: Geographic Related Information Display System (GRID)

Comments: GRID is a computer mapping program, first developed by Harvard University, used to map uniform data onto a line printer. The system is supported by the Southern California Regional Information Study (SCRIS) for the Census Use Study. GRID has three mapping options for the display of data: intensity, shading and value maps. The shading option prints one of ten possible shades in each cell depending on the data value for that cell. The density option prints in a cell the number of characters proportionate to the data value for that particular cell. The value option prints within the cell the numeric value of as many as two data items per cell.

GRID is coded in ASA standard FORTRAN so that it may be transported to almost any computer system. Included in the GRID system is a computer mapping language MAPTRAN. The design objectivity of GRID provides a system that can print simple maps quickly and easily with little preparation by a nonprogramming user and yet provide complex mapping capabilities for the sophisticated user.

Funding Agency: U. S. Department of Commerce, Bureau of the Census

Contact: Ronald E. Crellin, Project Manager, SCRIS, 5300 Santa Monica Blvd., Los Angeles, California 90029

References: Narojaro, "GRID Related Information Display System (GRIDS): Key to Instant Mapping of Local Census Data." Paper presented at the American Statistical Association, Detroit, Michigan, December 29, 1970.

Organization: Urban Data Center, University of Washington

Module: SACS (Street Address Conversion System)

Comments: SACS is similar to the ADMATCH System established by the Bureau of the Census. SACS was designed for geographic accuracy for plotting and measurement purposes rather than the identification of street addresses and census tracts for the purpose of mailing out questionnaires. As a result, the system utilizes far more accurate engineering diagrams than the blowups of USGS topological maps used by the Bureau of the Census to develop their address coding guide and ADMATCH system.

The system is designed around small machine use and is implemented on an IBM 1130, with access to a Burroughs B5500 provided for performing sorts on raw data from digitizers.

Funding Agency: The National Science Foundation

Contact: Charles E. Barb, Jr., Staff Assistant, Urban Data Center, University of Washington, Seattle, Washington.

Reference: Barb, C. E., Jr., "Street Address Conversion System", Proceedings of the 6th URISA, 1968, pp. 228-251.

Organization: U. S. Department of Agriculture Forest Service

Module: Map Information Assembly and Display System (MIADS)

Comments: The MIADS System is a punch-card based system. A grid is laid over a base map that measures 1/5 of an inch horizontally and 1/6 of an inch vertically. This corresponds to the space occupied by 2-print characters on a normal pica printer. A human interpreter then places two-digit codes into each of the 1/5 inch by 1/6 inch cells. These are transcribed onto the punch cards. (It is possible to insert character codes, but the system was intended to use numeric codes.) These cards form a base file which is manipulated by a program on an IBM 7090. The program will prepare a new set of cards from the base cards by selecting certain code cells and punching a new card deck which contains only those code cells in it plus the codes contained in the original cells. The card deck is then listed on a line printer and the line printer listing is pasted together to form a map of the same scale as that of the original base map.

Funding Agency: U. S. Forest Service

Contact: Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California

References: Amison, E. L., "A Computer Oriented System for Assembling and Displaying Land Management Information." U. S. Forest Service Research Paper PSW 17, 1964.

Storey, T. G., et. al., "INFORMAP--A Computerized Information System for Fire Planning and Fire Control." USDA Forest Service Research Paper PSW 17, 1969.

Organization: State of Alaska, in cooperation with Daniel, Mann and Johnson and Mendenhall

Study Group: A study was performed by Daniel, Mann, Johnson, and Mendenhall to establish a comprehensive land use plan for development, conservation, and social growth in the State of Alaska. It suggested that a land policy computer board model be developed to interrelate relevant factors influencing patterns of land use. This would try to simulate the consequences of alternate land use and development policies. A geographic information retrieval system is suggested as a necessary part to implement the model.

Funding Agency: Alaska Department of Planning and Research with U.S. Department of Housing and Urban Development

Contact: Daniel, Mann, Johnson and Mendenhall, West Division, 370 Petock Block, Portland, Oregon 97205.

Reference: Daniel, Mann, Johnson and Mendenhall, "A Comprehensive Land Use Plan for Development, Conservation, and Social Growth in the State of Alaska," November, 1970.

Organization: American Bar Association Committee on the Improvement of Land Title Records

System: CULDATA - Comprehensive Unified Land Data System

Comments: CULDATA is being developed by the American Bar Association to improve the access to land title records across the nation. The system involves uniform descriptions of land parcels as well as multiple indices to land parcels which allow access by parcel number, by geographic location, or by owner. The primary aim of this study group is to establish a legally acceptable nationwide land description. CULDATA is also known in Europe as the Electronic Cadastral System. The University of Cincinnati and the Economic Research Service of the U. S. Department of Agriculture have been the principal investigators of this system.

Funding Agency: American Bar Association and the U.S. Department of Agriculture

Contacts: Robert N. Cook, Professor of Law, University of Cincinnati, Robert T. Howe, Professor of Civil Engineering, University of Cincinnati Fred J. Lundberg, Director, Urban Data Center, University of Cincinnati

References: Cook, R. N., and Kennedy, J. L., eds. Proceedings of the Tri-State Conference on a Comprehensive Unified Land Data System (CULDATA). Cincinnati, December 9-10, 1966.

Howe, Robert T., Fundamentals of a Modern System of Land Parcel Records. University of Cincinnati, May 1968.

Organization: The American Institute of Planners Information Systems Department

Study Group: The Information Systems Department is a continuing study group which has not yet produced a report.

The Information Systems Department, created in the fall of 1970, is a division of the American Institute of Planners, the national organization of professional, urban, and regional planners in America. The department serves the various interests of planners, including:

Information systems: Collection, storage and retrieval of data and information; computerized processing of planning data.

Data series: The programs of the U.S. Bureau of the Census, land use inventories, employment, and other data concerning local and regional economies.

Planned design and evaluation: Demographic and econometric models, systems analysis, operations research, planning systems models.

Funding Agency: The American Institute of Planners

Contact: Donald D. Lamb, Chief of Land Use Planning, Southeastern Council of Governments or Herman G. Berkman, Professor of Planning, new York University, c/o The American Institute of Planners, 917 15th Street, N. W., Washington, D. C. 20005. These two gentlemen have been nominated for chairman of the Information Systems Department.

Reference: Circulars describing the creation of the Information Systems Department are available from the American Institute of Planners.

Organization: Association of Bay Area Governments

Study Group: The BRISC/Bay Region Employment Information Support Center, work program designed for the Association of Bay Area Governments by System Development Corporation, will be responsible for coordinating regional data sources that are now fragmented, sporadic and duplicated. The Center's principal objective is providing information for comprehensive planning in the San Francisco Bay area. The Center will also provide information services to a number of agencies in the region.

Funding Agency: Association of Bay Area Government

Contact: Robert A. Totschek, Urban Systems Research, System Development Corporation, Santa Monica, California 90406.

References: R. A. Totschek, "An Outline of the Work Program Design for a Bay Region Planning Information Support Center.", 7th URISA Conference, 1969, pp. 429-444.

Organization: Arizona State University Center for the Study of Urban Systems

Study Group: The Center for the Study of Urban Systems is reviewing the development of an environmental urban information system for metropolitan Phoenix. The proposed system will contain data on real property and land use. In addition, facilities to do metropolitan area counting are proposed. The bulk of the data in the system will be collected from assessors and appraisers.

Contact: Dr. William S. Peters, Director, Center of the study of Urban Systems, Arizona State University, Arizona

Reference: Peters, W. S., "Toward an Environmental Information System for Metropolitan Phoenix." Arizona Business Bulletin, June, 1967, p. 162.

Organization: California Statewide Information Systems Study

Study Group: This study was performed by Lockheed Missiles and Space Company to investigate and propose information systems for the State of California.

The study recommended a 10 year program with a total cost of $98.2 million to produce a computer based statewide information system. At the end of the third year a pilot program would be made available for testing at a total program cost of approximately $30 million. At the end of the fifth year, the first state agencies would begin to tie into the system at a total cost of approximately $45 million. Following the 10 year development program, estimated annual operating costs were $13.4 million.

This report, along with three others in related areas, was reviewed by various groups. The final review noted that in general, the reports were enthusiastically accepted by the monitor groups and by all reviewing agencies. The final review by the Automatic Data Processing Advisory Committee included "ADPAC urges an extensive action program on the statewide information system. The proposal is one of the most feasible ways of achieving improved services ...". ADPAC further recommended to the Director of Finance that initial planning be budgeted for the fiscal year beginning July, 1966.

Funding Agency: State Department of Finance

References:

"California Statewide Information System Study Final Report." Lockheed Missiles Company, Lockheed Aircraft Corporation. July, 1965.

Donati, R. W., "A Statewide Information System." Conference on Urban Planning Information Systems and Programs, 1966, p. 1.

"The Four Aero-Space Contracts: The Review of the California Experience." Department of Finance, State of California, Sacramento, January, 1966.

Organization: California Regional Land Use Information System Project

Study Group: The purpose of this study was to refine the concepts of the statewide information system proposed by Lockheed Corporation for the State of California to be used for land use information. The final report recommended that a system be developed over a period of six years at a total cost of $2.9 million, exclusive of operational costs.

Funding Agency: California State Office of Planning With HUD 701 Funds

References: "California Regional Land Use Information System Project Second Interim Report." Systems Group, 1 Park, Redondo Beach, California. August, 1967.

"California Regional Land Use Information System Final Summary Report." TRW Systems Group, 1 Space Park, Redondo Beach, California.

Organization: Canadian Meteorological Service

Study Group: Since climate and geography are profoundly interrelated, climatologists are keenly interested in geographic data which may be better used to interpret or exploit this archival information. A geographical file containing information for every 15 minute quadrangle of latitude and longitude in Canada is proposed. Included in this file will be information on elevation, slope, ruggedness, roughness, surface materials, drainage, proximity to and nature of geographic barriers, and land use. Pilot projects are currently being undertaken using the geographic and climatological control portions of the file. Limitations of the output are being appraised with a view to developing predictors of the effects of geography on climate.

Contact: Gordon A. McKay, Canadian Meteorological Services of Canada

Reference: R. F. Tomlinson, Editor, Environmental Information Systems, Proceedings of the UNESCO/IGU First Symposium on Geographic Information Systems, Ottawa, Ontario, Spetember, 1970.

Organization: Chicago Area Transportation Study (CATS)

Study Group: CATS, one of the oldest metropolitan transportation study groups in the country, is particularly unusual in that it has remained fully staffed since 1958. Many members of the staff came from the Detroit study group, DATS, which is generally recognized as the first. Many well known transportation planners have worked with CATS at one time or another. CATS is recognized as the first attempt to develop and apply a rational model of metro-area transportation planning. Computers were used in the early 1960's as well as sophisticated simulation models. The Cartographatron recorded on tape and used to display travel patterns, was one of the first applications of computer graphics to be used for planning purposes.

Funding Agency: 75% of the operating budget is from federal sources, 12% from state sources, and 13% from local sources.

Contact: Garred Jones, Director, Chicago Area Transportation Study

Organization: Denver Regional Council of Governments in cooperation with Peat, Marwick Mitchell, and Company.

Study Group: The Denver Regional Council of Government engaged Peat, Marwick, Mitchell, and Company to report on a conceptual design for a regional information system for the Denver Regional area. The system was to be developed to meet the projected data requirements of the Council's comprehensive program over both the long and the short term commencing in 1970.

Funding Agency: The Department of Housing and Urban Development

Reference: Peat, Marwick, Mitchell, and Company, Design of a Regional Information System, Denver, Colorado.

<u>Organization:</u> Department of Geography, University of Jerusalem

<u>Study Group:</u> The Address Coding Guide (ACG), and the ADMATCH/DIME Systems, are used by the city of Jerusalem under the guidance of the Department of Geography at the University. A mapping from these files is possible but not easy. The city has implemented a grid system linked to the ACG system which provides better facilities for analysis and mapping. The SYMAP mapping program is used to print map outputs.

<u>References:</u>

Tomlinson, R. F., Environmental Information Systems, Proceedings of the UNESCO/IGU First Symposium on the Geographic Information Systems, Ottawa, October 1970. Publication of the International Geographical Union.

Shahar, Arie, "Mapping of Jerusalem by Computer," Computers and Automation. Ottawa, May 1970.

<u>Organization:</u> Indiana Office of Traffic Safety

<u>Study Groups:</u> The State of Indiana has constructed a promising framework for a unified and effective traffic control and accident prevention system. The Indiana Traffic Accident System (INTRAC) is actually a large geographically oriented data base on traffic accidents. The system is composed of a series of programs to edit and logically organize this data base. Because of the specific nature of INTRAC, it has little or no use in the State of Illinois. The success of the use of this file has led to a newly initiated study of an entire information retrieval system for the State of Indiana.

<u>References:</u>

John E. Stoner, Phil E. Johnson, "The Possibilities of Integrating Data Processing for the State of Indiana and its Political Subdivisions. Based upon traffic safety information system developed for the State Highway Commission", available from Data Indexing Systems Corp., February 14, 1969.

Organization: International Geographical Union (IGU) Commission on Geographical Data Sensing and Processing.

Study Group: The primary objective is to "... explore the scope for Environmental Data Handling in digital form and (determine) how it may be used to shorten the period between the collection of data and their utilization." The first symposium on geographical information systems was held in Ottawa, Canada between September 28 and October 12, 1970. Meetings were held under the joint auspices of the Natural Resources Division of UNESCO, the Canadian National Commission, and the Commission on Geographical Data Sensing and Processing of the International Geographic Union.

Contact: R. F. Tomlinson, Office of the Secretariat, Cabinet Committee on Planning and Programs, Government of Nova Scotia, P. O. Box 505, Halifax, Nova Scotia.

References: Tomlinson, R. F., Environmental Information Systems, Proceedings of the UNESCO/IGU First Symposium on Geographical Information Systems, Ottawa, September 1970. Publication of the International Geographic Union, Commission on Geographic Data Sensing and Processing.

Organization: Los Angeles Planning Department

Study Group: The department is developing the Los Angeles Municipal Information System (LAMIS). In 1963 the mayor and city council approved an ordinance providing for the centralization of data processing systems within the Data Service Bureau. No information about the present condition of this proposed system has been published.

Reference: Johnson, Land O., "An Automated Data System-the Los Angeles Fourth Conference of Urban Planning Information System and Programs.

Organization: Metropolitan Washington, Council of Governments, District of Columbia.

Study Group: The Council of Governments has developed techniques for storing current transportation data such as origin destination surveys and land use data. Their system includes use of computerized addressedcoding guides, based upon the census addressed coding guide, to locate each residential address to the nearest census block.

Contact: J. C. Barrett, Director, Data Systems, Metropolitan Washington, Council of Governments, Washington, D.C.

References:

J. C. Barrett, "Structuring Regional Data." 7th URISA Conf., 1969. pp. 74-91.

G.V. Wickstrom and A. E. Pisariski, "The Use of Longitudinal Information Systems for 'Continuing' Urban Transportation Planning." 7th URISA, 1969 pp. 382-393.

Organization: National Aeronautics and Space Administration, Mississippi Test Facility

Study Group: Jack Balch, the director of the Mississippi Test Facility is establishing a Regional Environmental Center for Arkansas, Louisiana, Mississippi, and possibly Alabama. Bruno Sables is surveying geographic information systems.

Funding Agency: National Aeronautics and Space Administration

References: Telephone conversations. No publications have been produced.

Organization: The President's Commission on Statistics

Study Group: This commission has interests in national geocoding programs for statistical data, specifically small area statistics, and has cooperated with the Census Bureau, with the Committee on Small Area Statistics of the American Statistical Association, and with census data experts at workshops which it has sponsored.

Funding Agency: United States Government

Contact: John W. Lehman, Executive Director, American Statistical Association, 806 15th Street N.W., Washington, D.C. 20005

Organization: The President's Council on Environmental Quality (CEQ)

Study Group: CEQ has a contract with Mitre Corporation to investigate the Nature of environmental indicators for air, water, and land pollution. The final report produced by Mitre has not yet been released; however, the CEQ did make available to the Center for Advanced Computation some copies of the report and the appendices to their study. The purpose of the study was to develop a national environmental monitoring system. The report describes a basic taxonomy broken down into air, water, and land classifications within which are data pertaining to particular physical, biological socio-economic, cultural and aesthetic factors. For example, the level of lead concentration in the air might appear as an index which is classed as a physical parameter in air data. In total, 112 indices were developed.

Funding Agency: Presidents Council on Environmental Quality

Contact: Richard S. Greeley, Associate Technical Director, Systems Development Division, MITRE Corporation, 1820 Dolly Madison Blvd., McLean, Va.

References: "Monitoring the Environment of the Nation." MITRE MTR-1660. April 1971.

"Monitoring the Environment of the Nation-interim Report--Legislative Proposals and Budget Estimates." MITRE MTR-4173 October, 1970.

Organization: Southern California Association of Governments

Study Group: The TRW Systems Group performed a preliminary design for a regional information system as an integral part of the statewide system for the Southern California Association of Governments. This system is more a government information system than a geographic information system and is designed to integrate into the Lockheed state-wide information system proposal.

Funding Agency: Southern California Association of Governments

Contact: Thomas E. Sawyer, Intern in Public Affairs, Coro Foundation, TRW Systems Group, 1 Space Park, Redondo Beach, California.

Reference: "Preliminary Design for Regional Information Systems as an Integral Part of the State-wide System." TRW Systems Group, August 1967.

Organization: The Office of Science and Technology, Executive Office of the President, Committee for the Study of Environmental Quality Information Programs in the Federal Government (CEQUIP).

Study Group: The executive Office of the President has established an ad hoc group under the chairmanship of Dr. Henry Kissman of the FDA to study information programs related to the enhancement of the quality of the environment. As a first step, this committee will try to identify and study all those information programs which support environmental pollution missions. A work shop was scheduled for May 18 and 19th, 1970 to discuss the contents of their report which the committee expects from the various information programs, to consider problem areas, to ascertain that all information programs open to the study have been recognized, and lastly, to identify operational techniques and procedures of potential common interests. A final report is due to be released shortly.

Contact: Henry M. Kissman, Chairman, CEQUIP Committee, Room 3001, 200 "C" Street SW, Washington, D.C. 20204

Reference: Environmental Data Bank, Hearings before the Subcommittee on Fisheries and Wild Life Conservation of the Committee on Merchant Marine and Fisheries, House of Representatives, 91st Congress, Second Session.

Organization: United States Air Force Aeronautical Chart and Information Center

Study Group: The Aeronautical Chart and Information Center is developing a system to aid cartographers in the generation of aeronautical maps. It is of little or no use in a general geographic information system.

Funding Agency: United States Air Force

Contact: Mr. Arthur L. Ziegler, Aeronautical Chart and Information Center, St. Louis, Missouri 63228

Reference: Cross, Benjamin Joseph and Cade, Jaunetha N., "Cartomation" ACSM. March 7-12, 1971.

Organization: U.S. Department of Commerce, Bureau of the Census

Study Group: The Census Use Study, a small area data research study, sponsored by the Bureau of the Census was established in New Haven, Connecticut in September 1966 and concluded in July 1969. It was established to explore the current uses and future needs for small area data handling and display techniques in local, state, and federal agencies. In response to established goals of the study, exhaustive research was carried out in the following areas: geographic based information systems, record matching, computer mapping, special tabulations of data, special sample surveys of family health and area travel patterns, and local data user's interests and needs.

Funding Agency: Department of Commerce, Office of Civil Defense of the Department of the Army, Department of Health, Education and Welfare, Department of Housing and Urban Development, and the Department of Transportation.

References: Census Use Study-report #1-general description, U.S. Department of Commerce Bureau of the Census. Other reports are available from the publications distribution section, Bureau of the Census, Washington, D.C. 20233. Commerce Department.

Organization: U.S. Department of the Interior, Geological Survey

Study Group: The U.S. Geological Survey (USGS) is participating in several remote sensing projects. The CARETS (Central Atlantic Regional Ecological Test Site) project is directed by the USGS in cooperation with the National Aeronautics and Space Administration and with the participation of other Federal, state and local agencies having environmental interests in the region. Included in the projects are such study "subsites" as the Chesapeake Bay, Delaware Bay, the coastal Islands, and the Washington-Baltimore metropolitan area. The Geological Survey is also interested in similar projects in the Boston area and the San Francisco area.

The study uses high altitude air craft and will eventually use satellite data from ERTS (Earth Resources Technology Satellite).

Funding Agency: The U.S. Geological Survey and the National Aeronautics and Space Administration.

Contact: Dr. Arch Gerlach, USGS Chief Geographer, U.S. Geological Survey, 1340 Old Chain Bridge Road, McLean, Va. 22101

References:

Simpson, R.B., "Production of High Altitude Land Use Map and Data Base for Boston." Final technical report on Phase I of Contract #14-08-0001-12640. Geographic Applications Program, Geological Survey, U.S. Department of the Interior, Washington, D.C.

Wray, J.R., "Census Cities Project and Atlas of Urban and Regional Change." Paper presented at the International Workshop on Earth Resources Survey Systems at the University of Michigan, Ann Arbor, Michigan, May 3-14, 1971.

"Central Atlantic Region to be Studied From Air and Space." News release for June 11, 1971.

"Additional "3-D" Urban Environment Part of Studies Plan." News release for October 28, 1970. U.S. Department of the Interior Geological Survey.

Organization: U.S. Department of Housing and Urban Development, Federal Urban Information Systems Interagency Committee (USAC)

Study Group: Nine federal agencies from the urban information system inter-agency committee (USAC) headed by the U.S. Department of Housing and Urban Development are sponsoring the development of prototype municipal information systems in six U.S. cities. These cities are Charlotte, Dayton, Long Beach, Reading, St. Paul, and Wichita Falls. A report on urban information systems is published quarterly as review of the USAC program. Many of these systems are described in this document.

References:

Report-Urban Information Systems, U.S. Department of Housing and Urban Development.

Organization: Coordinated Science Laboratory, University of Illinois

Study Group: The Kankakee Project was a study in urban growth. Kankakee, Illinois, a residential city with a population of approximately 50,000, was chosen as a center for the study because of it availability of data. In particular, complete records were available on real estate transactions, construction, and land use over a period extending from 1854 to 1969. Extensive computer graphic facilities were developed for this project to display data and make the on-going processes both understandable and available to the researchers.

Funding Agency: Ford Motor Company's Joint Services Electronics Program and the National Science Foundation.

Contact: W. J. Bouknight, Center for Advanced Computation, University of Illinois, Urbana, IL. 61801

References: L. P. Kadanoff, J.R. Voss, and W. J. Bouknight, "A City Grows Before Your Eyes," Computer Decisions. December 1969, pp. 16-23.

Organization: Urban and Regional Information Systems Association (URISA)

Study Group: URISA has established a special interest group in geographic base file systems and has presented many symposia on urban and regional information systems. Proceedings are readily available and of highly variable value.

Contact: Charles E. Barb, Jr., Asst. Director Urban Data Center, 125 More Hall, University of Washington, Seattle, Washington, 98105.

Organization: United States Air Force, Rome Air Development Center (RADC), Experimental Cartographic Facility

Module: Automated Photogrammetric System

Comment: RADC has developed an automated photogrammetric system to assist in the preparation of aeronautical charts and maps. It is of limited use in the preparation of geographic information system data.

Funding Agency: United States air Force

References: "The Evolution of Automation and Cartography at Rome Air Development Center." Air Force Systems Command. Griffiss AFB. March 1970.

"Development of Raster Data Handling Techniques at Rome Air Development Center." Air Force Systems Command. Griffiss AFB. March 1971.

Organization: R. L. Polk and Company

System: Polk Urban Information System

Comments: The Polk System is designed to translate data into meaningful statistics grouped by small areas and presented either in tabular or graphic output. We do not know whether this system is available to outsiders as a product, or if it is for internal use only.

References: Blatt, Donald H., "Applications of the Polk Computerized Urban Information System to Urban Economic Planning." Proceedings of the Sixth URISA Conference 1968.

Organization: Bureau of State Planning, Wisconsin Department of administration

Study Group: A series of recommendations for the development of an information system were made in a report to the Bureau of State Planning. One of the essential recommendations was that the information system must have a spatial data manipulation capability. System capabilities must be developed 1) for spatially identifying data, 2) for retrieving data by spatial and other criteria, and 3) for graphically displayed data.

References: Kenneth J. Dueker, A Recommendation for a State Planning Information System, A Report to the Bureau of State Planning State of Wisconsin, August 1967.

APPENDIX C

FEATURES AND CAPABILITIES AVAILABLE IN EXISTING GEOGRAPHIC INFORMATION SYSTEMS

Nature of the Data

Natural resource and socio-economic data have one common characteristic--a geographic location. The data vary in their geographic representations. Effluent sources and special points of interest are frequently represented as attributes of a single point, streams and roads as integrated point and line data which form networks, and economic and natural resource data as attributes of a specific area. Although geographic information systems and their components address one or more data representations, only recently have systems been planned to address all three.

Geographic Referencing Schemes

Several geographic referencing schemes are used to manipulate uniform grid planning data. Point and line data manipulation normally use Universal Transverse Mercator, latitude and longitude geodetic coordinates, and sometimes an arbitrary x,y coordinate system which is overlayed on the geographic area to be analyzed.

Systems which record data on areas often use State Plane, Universal Transverse Mercator, or the rectangular land survey coordinates. Rectangular survey referencing specifies an area rather than a specific geographic point. Then, by indicating exact distances in terms of feet from an east or a west and a north or a south boundary of the area containing the point, geographic points are further specified. Because of survey errors, using rectangular survey coordinates creates a small percentage of irregularly shaped areas, i.e. non-uniform grid cells.

Some systems reference areas by a single point located within the area. In those cases, it is common to calculate, either exactly or with a rough visual approximation, the centroid of the particular land area involved. All data in that area are then assumed to be at the centroid of the cell for coordinating referencing purposes.

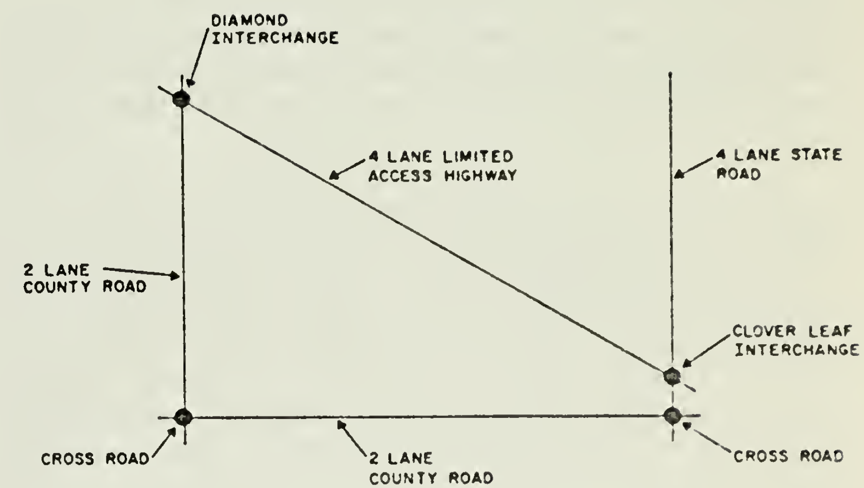
Point, Network, and Area Data

The data structures used in geographic information systems mirror the point, line, and area nature of the data.

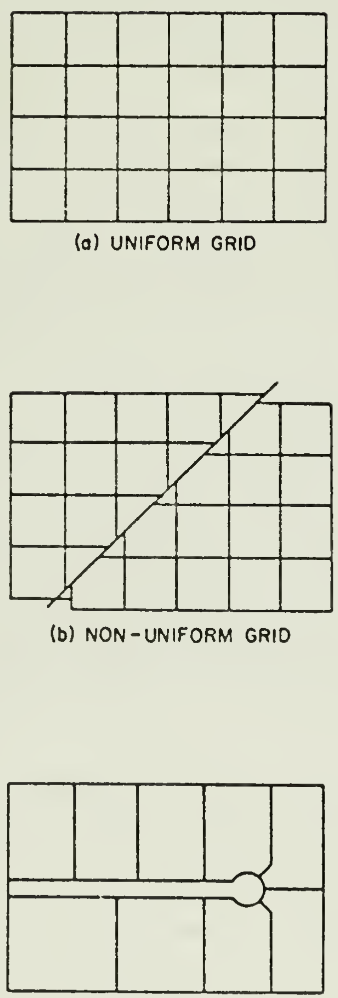
Typically, node and link structures appear in network systems. These systems may store data on transportation or stream networks (see Figure 1). Depending upon the kind of analyses performed, data may be stored on network nodes, links, or both. For example, to determine the load-carrying capacity of a road network, the data normally used are a property of the network links. Alternatively, when examining network access in a transportation system, the network is viewed as a node system in which the most vital data are at the entrance and exist nodes of the system. Hydrologic and stream models often use node and link data attributes simultaneously. In this case, gauging or sampling stations are nodes on the network and parameters such as assimilative capacity or time of travel are attributes of the network links.

Area geographic structures are found in three basic configurations: uniform grid, parcel, and area boundary systems.

Uniform grid systems superimpose a square or rectangular grid over an existing land area (see Figure 2a). These systems have the advantage of simplified software and information retrieval algorithms. However, the major costs of the geographic information system involve data acquisition and entry. Little data exists on any uniform grid. Thus, acquisition effort is normally required for grids and frequently over shadows any gains made in easy computer system development.



ROAD NETWORK
FIGURE 1



(a) UNIFORM GRID

(b) NON-UNIFORM GRID

(c) GENERALIZED PARCELS

PARCEL SYSTEM CONFIGURATIONS

FIGURE 2

Furthermore, a grid size for these systems must usually be chosen before data collection begins. If the grid is very small, overhead will occur in monitoring of the large number of cells needed to cover a region. If the grid is large, the resolution of data for urban regions, where land use information is extensive, might be too low to be useful.

In parcel systems, geographic regions are divided into non-overlapping cells whose union is the entire region. Data is then collected from the geographic area represented by a particular cell and stored, usually as one record, within the system. Parcel systems are further categorized into non-uniform grid, and generalized parcel systems.

Non-uniform grid systems are being developed because large quantities of data have already been associated with the legal land description of the Rectangular Survey Coordinates (c.f. Figure 2b). (This survey system was originally intended to be a uniform grid system. Unfortunately, due to curvature of the earth, irregularities have occurred. Furthermore, indirect survey errors, caused by Indian boundaries which surveyors could not cross until a treaty was signed, resulted in several years' delay in completing the survey. This forced surveyors to begin again at a new base line which did not match up with the old one. As a result, approximately 2% of all sections in the State of Illinois are irregularly shaped. The nearly uniform grid nature of the system allows the user to address the cell 98% of the time as if they were squares on a checkerboard. However, the 2% irregularities require computer systems to treat all cells internally as if they were parcels of arbitrary size and shape.

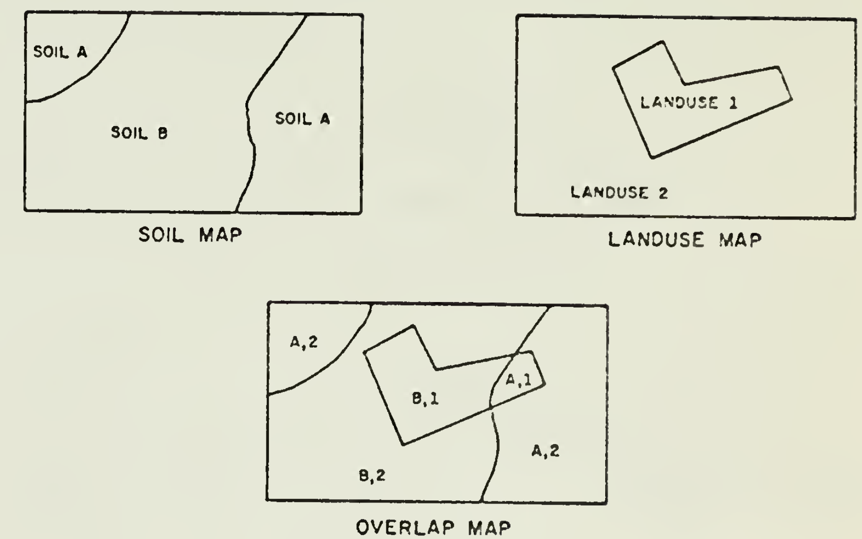
Generalized parcel systems make no prior assumptions about the shape of any parcels; parcels are usually described as arbitrary polygons (c.f. Figure 2c). In the past, these systems have been developed to store land ownership data; presently, they are being explored for potential storage of generalized environmental information.

Although the multiple resolution problem described previously for a uniform grid system can be solved using generalized parcels, digitizers or scanners are required for inputting parcel boundaries, and an easily used parcel addressing scheme is difficult to specify.

An area boundary system stores a digitized polygon of an actual geographic boundary. These systems are used to store images of maps. Usually, several maps are used to describe a region with one attribute described on each. A single value of an attribute is contained within a boundary (c.f. Figure 3). These systems provide capability for overlapping and deleting collections of areas and their associated boundaries.

Area boundary systems are the most general and potentially the most accurate representations of real world data, but these systems have significant disadvantages over previous systems described. Operational costs are high because excessive processing time and large memories are required. The number of overlays which may be done simultaneously tends to be low because of overlayed map density. Despite such disadvantages, these systems may be useful for creating a parcel system data base by overlapping maps of parcel boundaries and other mapped data; in this case, the area boundary system is used infrequently and only as an archival data base.

The planner, searching for trends in resource utilization, has a need not for detailed data but for aggregated data of a higher level than the area boundary of a plot of trees or a single plot of soil. Unfortunately, if one collects data on a parcel system basis or on a uniform grid system basis and then decides to change his parcel or grid size, he may require a new data collection effort to reaggregate the data into new parcels or grid cells. If the data has been originally collected and stored in an area boundary system, a computer program can be written which will automatically extract the original low level collection effort from the area boundary system and aggregate it into



AREA BOUNDARY SYSTEM
FIGURE 3

the cells or parcels desired for the uniform grid or parcel system. The cost and analysis advantages of uniform grid and parcel systems for data manipulation can then be exploited.

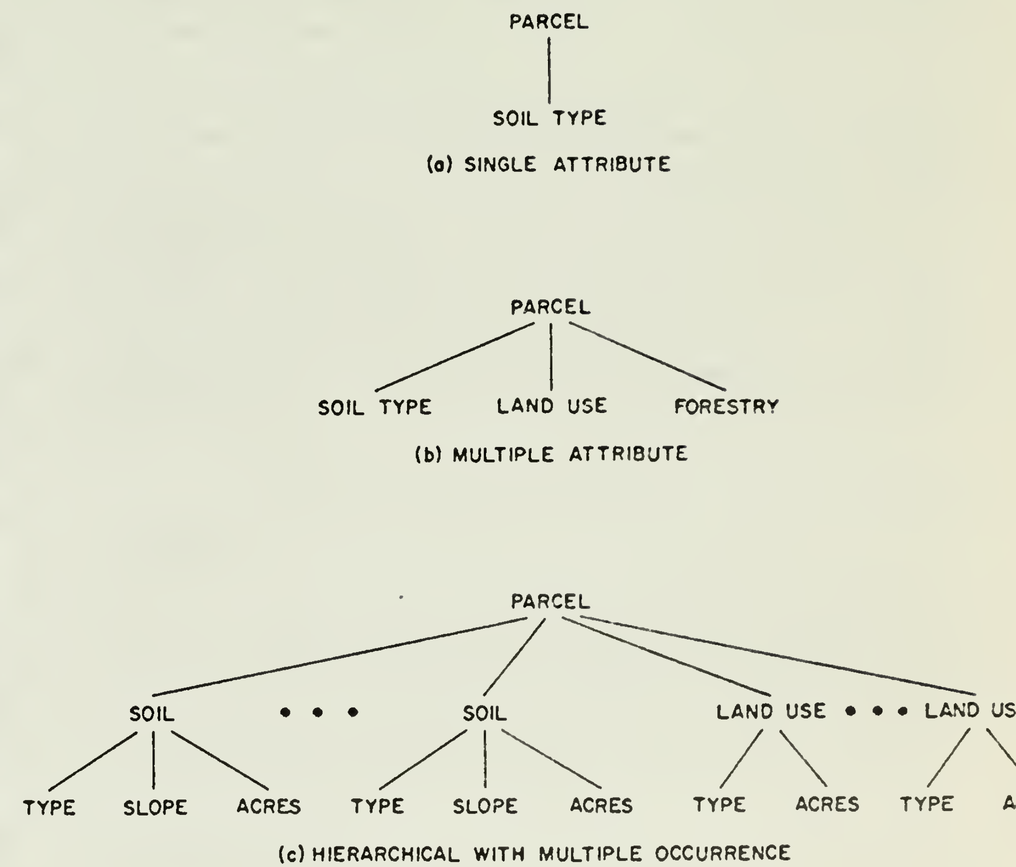
Data Structures Used in Geographic Information Systems

Data are stored according to a fixed structure for each parcel, area boundary, line, or point in a geographic information system. Retrieval capabilities depend on the complexity of the data structure. At the simplest level there are single attribute data structures (see Figure 4a). For instance, in some systems each uniform grid cell can have only a single attribute with a single associated value which can later be mapped out.

On the other hand, most systems allow several attributes and their associated data values for each data cell (see Figure 4b). These attributes are referenced by name and can each have one of several numerical or alphabetic values. Attributes in a single cell might include the linear footage of a stream and types of forest cover.

Trees or hierarchies represent the next level of complexity for data structures (see Figure 4c). In these structures only the leaves, that is, the nodes which do not have any further branching, represent real data elements. All other nodes represent groups of data elements. Hierarchies permit referencing several values of the same data item by one name. For example, in a two-level system, the first level may have class names - soils, hydrology, and land use. Then, under soils might be data items - type, acres, slope, and erosion all of which can be referenced at once by referring to "soil data."

If a hierarchical structure also allows several occurrences to be stored for each class, every observation for a data item for that class can be stored rather than just one value for the entire parcel. For example, data could be stored separately for each forest plot. If



DATA STRUCTURES
FIGURE 4

sub parcel location for each plot of trees within the parcel is stored, higher data resolution than the geographical referencing resolution of the parcel is added. Thus the data on each plot in a land parcel will be stored along with values of other attributes and can be referred to under the general class "forestry." Each occurrence of forestry data will then have the same structure as the others, but different data values. Another advantage of multiple occurrences or observations is that selected occurrences can be retrieved only if conditions are met on the associated data items. A particular parcel can be retrieved, for example, only if it has a plot of soil with a particular soil number that also has a certain erosion factor in the same soil plot.

Although two-level trees are adequate for natural resource data, social and economic data may require trees containing more levels. Data structures more complex than trees for a single parcel or area boundary are not currently used in geographic information systems.

Data Input

A variety of techniques are used to input data into geographic information systems. A common technique uses overlays, which outline the parcels, to input data on maps. The overlay is placed on the data contained within the boundaries of each parcel in the overlay are manually recorded.

Many experimental and several established systems use sophisticated hardware. Technology involving digitizers and scanners has been proved efficient, reliable, and accurate, and is used in almost all of the cartographic systems investigated. When using a digitizer, the data encoder moves a special device to point to different areas or boundaries on a map. Digitizer hardware is capable of recognizing to a very high degree of accuracy (about .004 inches) the location of the pointing device over the map. When requested by the operator, this location is recorded on tape and, depending upon the number of options

on the digitizer, data is automatically entered from a deck of punched cards, at a keyboard, or through a variety of other techniques to describe what information is to be recorded along with the coordinates of that particular point.

Digitizers are used for several purposes. They can be used to initially describe the actual boundaries of areas, states, counties, or ownership parcels. They can also be used to point within a particular parcel or grid cell. In the latter case, the digitized point is used as a reference to identify to which cell data is to be associated. For instance, a parcel map may be drawn and digitized to produce computer usable boundaries. Then a parcel numbering scheme is devised. In a second digitizing step the operator points to each parcel in numbered sequence. The computer computes which parcel the operator is pointing to and records the next number in sequence as belonging to that parcel.

Scanners can be used to automatically recognize and digitize lines on a map without requiring human operators to follow the lines manually.

Another input problem is presented by data that already have some kind of urban geographic locator, usually street address or zip code. Special computer programs are written to convert these locators into cell or parcel locations for storing data within the system. Several programs have been written to convert street addresses into census tract locations.

Storage Techniques

One of the most important abilities for geographic information systems, especially for those systems that operate in conversational mode, is the ability to compress data. Software modules should be designed to accommodate the automatic reduction of the number of data bits required for each data element value. These compression processes

can be executed without the user's knowledge. For example, the data element soil slope may have one of seven values for a plot of soil: A, B, C, D, E, F, or G. Eight data bits are required to store this data as a letter in the computer, but because only seven values are to be represented, the system only needs to store three bits in the data base. When, for example, a user requests the amount of soil acres of a particular parcel in which the soil slope is A, the request is translated to ask only for a bit pattern of three bits to be looked up in the data base. When the results are printed, the bit patterns stored in the data base are automatically converted into a letter.

Whether the individual files and records are to be accessed sequentially or randomly must also be determined. Sequential access requires only that a system be able to get at the next record after processing a previous record (typical of tape files). However, in direct accessing techniques, several internal reference tables must be stored which record the addresses or particular parcels within the system based on geographic locations associated with data records or data attribute values stored in data records.

Types of Analyses Performed

Most geographic information systems provide some form of tabular summary capability and some ability to do sorting. Sophisticated tabular summary systems prepare reports which aggregate data up to larger geographic areas. If the system contains data on census blocks, it allows aggregation up to a census tract or a block group for reporting purposes. In addition, these systems may provide data distribution descriptions in tabular summaries, thus providing further breakdowns, such as tract subtotals for all data items.

As more manipulation capability is provided, arithmetic functions can be described to give weight to various attributes of the cell according to the environmenta for a specific project, a highway or

airport location, for example. Once each cell is appropriately weighted, it is almost impossible to examine the impact of this weighting choice on a large geographic area without some graphic facility. Thus, it is common to find a sophisticated graphic facility which not only draws basic boundary maps, but also shades cells in the maps. If high positive weights in cells correspond to a high suitability for siting an airport, a gray scale map graphically displaying these weights would indicate that darkest areas are best locations, lightest are worst.

Some geographic information systems allow limited modeling capabilities for specific problem areas. These modeling efforts fall into three basic classes, simulation models, mathematical models, and statistical models. Simulation models are highly specific, and are used primarily in very complex situations where adequate analytic mathematical models for describing the system do not exist. Simulation packages are found most often in urban area systems.

Mathematical models such as linear programming or systems of partial differential equations have been used. A very limited amount of work has used linear programming systems, but several hydrologic differential equation models have been generated for problems based on specific river basins and stream networks.

Statistical models are widely used. Because the amount of available economic and demographic data is large, the most common method of manipulating this data is integrating statistical routines into the system or providing output files to be input to separate statistical systems, usually found on the same machine.

Techniques must be developed to limit the area of data search and retrieval.

Manipulation Capabilities

Describing study regions is common. Study regions are frequently created or described simply by indicating the coordinates or lines which bound the particular study area. All systems with the ability to describe study areas have this facility. More sophisticated systems also provide an ability to describe the study areas in terms of cell content; that is, cell content can determine whether or not a cell should be included in a study area.

If a forester is concerned only with those areas in a county which contain a red pine forest cover, he can create a study region which consists only of those cells which contain red pine. An interesting feature of content-defined regions is that they can also describe discontiguous regions, the red pine in a county, for example. Although capabilities for retrieving data from or merging, intersecting, or excluding discontiguous regions are not usually provided, the following capabilities are shared by many systems:

1. statistical data manipulation

2. data retrieval according to boolean functions (i.e., conditional retrieval with respect to the operations of AND, OR, and NOT)

3. data analysis using arithmetic functions

4. special graphic analysis

Some systems also permit attaching arithmetic values to character data. In a hypothetical environmental impact analysis, it might be important to attache some numerical weight to a soil described in character format, such as soil type W103. Almost all of the systems investigated have some capability to do minor mathematical manipulation, but only a

few of the more sophisticated can handle general arithemetic, Boolean, or character manipulations without introducing special purpose packages into the system.

Access Techniques

The mode in which a user accesses his information can be extremely important. In batch systems a user submits a request, usually on cards, and waits an hour or several days before he gets a response. This mode of operation prohibits "hunch" or exploratory work by users of geographic information data bases. In contrast, conversational systems allow an analyst on-line access to his data. He can seek areas in which to do detailed analyses, he has a better intuitive feel for the actual environment he is studying, and the time he must spend to solve a particular problem is significantly reduced.

The type of languages used to interrogate a particular geographic information system must also be considered. At the very least, most systems provide, within an existing programming language such as FORTRAN or PL/1, some subroutines or procedures to manipulate geographic information. These systems require experienced programmers to actually perform the data manipulations required for most studies.

At the next level of utility are canned programs for which a user has only to specify certain parameters. These languages tend to be very cryptic; their symbolic requests can be understood only by frequent reference to manuals. And, because of their rigid accessing language capability, these systems have limited manipulation capability.

The most sophisticated systems provide the user with a problem specific language designed especially for geographic information retrieval. These languages are complete, rather general, and allow within their structure all of the manipulations permitted by the system.

A major characteristic of these languages is the ability to reference data elements by their names using a text similar to English. This increases the readability of requests and decreases the amount of instruction required by a user.

Output Techniques

There are two basic output forms in geographic information systems: tabular reports and graphical displays. Most geographic information systems provide at least the graphics output package; only a very few provide general purpose tabular report generators.

Due to the specific nature of regional data coverage and the problem orientations of most geographic information systems, a special purpose or canned report generator is adequate. Systems with more powerful report generation capabilities allow the user to output his tables using his own specified formats and prepare aggregations and distributions of data for tabular presentation. Some systems generate files to be shifted to other systems (e.g., statistical systems or simulation packages on the same or on a different machine). This generation of a computer file for use on a system other than the original one is a kind of report generator; its advantage is the ability to permit data collected in one system to be formatted so that it is usable in another system, one which may not be a geographic information system at all. By using data already stored in other geographic information systems, data collection and recording funds are saved.

The second kind of output is a graphical display. Various graphic routines are available which draw points, lines, and polygons. A great many geographic information systems are actually little more than graphics packages which provide a nice presentation of pre-analyzed geographic information. These graphic packages are sometimes augmented to handle a small number of geographic data storage and manipulation requirements.

Line printer plots are fast and cheap, but they have extremely low resolution. Studies like the New Haven Census Use Study indicate that maps coming off line printers are inadequate and must usually be enhanced with manually drawn reference lines (e.g., roads and city boundaries).

High quality graphics can be provided by pen, CRT, or photo plotters. Plotters, which vary in accuracy from .01" to .0001", provide far more accuracy than is needed for planning purposes, but only the very highest accuracy of .001" or .0001" resolution is adequate for engineering drawings.

Another type of hard copy graphic display unit has recently appeared on the market, the electrostatic printer or plotter (e.g., AB Dick, Clevite, and Gould units) which provides .01" resolution and is very fast and inexpensive. Images are produced in one of two ways, by spraying charged ink particles and deflecting them to an appropriate spot on the page, or by charging specially coated electrostatic paper and depositing charged graphite or some other kind of toner on the surface as the charged areas pass beneath the toner head. These units have a price, resolution, and capability compatible with the needs and budgets of local and regional agencies.

General Purpose Information Management Systems

Several general purpose information management systems exist which permit the user to specify data elements by name, to input data through terminals, card readers, or tapes, to and store and access the information. They can be used as geographic information systems if some of the attributes stored with each record in the system are geographic coordinates. But since geographic referencing is not an inherent system capability, a user must explicitly specify coordinates upon which retrieval is to take place.

Data compression facilities are essential to handling large amounts of data, especially when used in an interactive mode. Generalized information retrieval systems do not provide this automatic compression facility; a user must specify all data in special coded formats to reduce storage requirements. Rather than specifying a value of a data element in what might most naturally be character format, he must specify the value of that item in terms of integers that can be directly translated into compressed bit patterns. Since data compression facilities are not provided, access and storage of information are more expensive than in a tailored geographic information system, and manual translations of data values are necessary. Geographic information systems without proper analysis tools perform only half the job. Although generalized systems provide excellent report generators, very little attention has been given to graphic output. Sometimes the only way in which information can be properly displayed for regional planning is maps. The system's complexity makes mapping integration unlikely without complete redesign. Nor do these systems provide the extensive analysis aids necessary to environmental planning.

Still another problem with many of these systems is their use of particular kinds of language interface. Systems can be classified according to the way in which a user must specify his requests. "Extended programming language" systems use extensions to programming languages such as COBOL and PL/1, along with subroutines and added language constructs that relate to the creation and manipulation of data structures. These systems are not user-oriented and would be very difficult for a planner to use; but they might be used by computer scientists as a foundation on which a geographic information system could be built.

The second type of language interface is a "form controlled" system. Because these are not on-line systems, they have many drawbacks. A user makes a request by completing a form which system

operators translate into subroutine calls to retrieve, analyze, and generate reports. Because these systems are off-line, they preclude effective "hunch" or exploratory data searches by regional planners.

The third type of language interface is provided by systems with their own interface language, systems which are typically time-shared or remote entry multi-programmed. These provide a good user-oriented language interface, but many of the spatial language constructs found necessary in geographic information systems are not provided.

Summary

There are many desirable capabilities available in existing geographic information systems. Unfortunately, these facilities are spread over many systems. Each system has only a fraction of the facilities needed by Illinois decision makers and planners. Furthermore, many of the advanced capabilities, while demonstratable and implemented in some systems, are often one shot attempts which cannot be integrated with other systems and should not be until they are cleaned up and made compatible with more general data structures and user language concepts.